DOCUMENT RESUME

ED 304 011 FL 017 808

AUTHOR	Powers, Donald E.
TITLE	A Survey of Academic Demands Related to Listening Skills. Test of English as a Foreign Language Research Reports Number 20.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-85-48
PUB DATE	85
NOTE	96p.
PUB TYPE	Reports - Research/Technical (143) -- Statistical Data (110) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; College Faculty; Comparative Analysis; Construct Validity; Educational Needs; *English (Second Language); Higher Education; Language Tests; Listening Comprehension; *Listening Skills; Questionnaires; Statistical Surveys; *Teacher Attitudes; *Test Items
IDENTIFIERS	*Test of English as a Foreign Language

ABSTRACT

Undergraduate English faculty and faculty in six graduate fields were surveyed about students' academic listening needs. The objectives were (1) to obtain perceptions of the importance to academic success of various listening skills and activities, (2) to assess the degree to which both native and non-native speakers experience difficulties with these skills or activities, and (3) to determine faculty views of alternative means of evaluating these skills. Results show faculty perceived some listening skills as more important than others for academic success, including nine skills related primarily to aspects of lecture content. Faculty perceived that non-native students experience more difficulty with all listening activities, and that non-native students have disproportionately greater difficulty with some activities, such as following lectures given at different speeds and comprehending or deducing the meaning of important vocabulary. With respect to measuring listening comprehension, some general approaches and specific item types were judged more appropriate than others, including tasks entailing recall of details and inference or deduction. The results are used to suggest further research on the construct validity of the listening comprehension section of the Test of English as a Foreign Language. The survey questionnaire is appended. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED304011

TOEFL

TEST OF ENGLISH AS A FOREIGN LANGUAGE

Research Reports

REPORT 29
DECEMBER 1988

A Survey of Academic Demands Related to Listening Skills

Donald E. Powers

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
WEIDENMILLER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL017808

ETS
EDUCATIONAL TESTING SERVICE

2

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of over thirty organizations, public and private, that were concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program and in 1973 a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations (GRE) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the GRE Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

A continuing program of research related to TOEFL is carried out under the direction of the TOEFL Research Committee. Its six members include representatives of the Policy Council, the TOEFL Committee of Examiners, and distinguished English-as-a-second-language specialists from the academic community. Currently the committee meets twice yearly to review and approve proposals for test-related research and to set guidelines for the entire scope of the TOEFL research program. Members of the Research Committee serve three-year terms at the invitation of the Policy Council; the chair of the committee serves on the Policy Council.

Because the studies are specific to the test and the testing program, most of the actual research is conducted by ETS staff rather than by outside researchers. However, many projects require the cooperation of other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL program office. Local research may sometimes require access to TOEFL data. In such cases, the program may provide this data following approval by the Research Committee. All TOEFL research projects must undergo appropriate ETS review to ascertain that the confidentiality of data will be protected.

Current (1984-85) members of the TOEFL Research Committee include the following:

Henry F. Holtzclaw, Jr. (chair)	University of Nebraska
Kathleen M. Bailey	Monterey Institute of International Studies
Paul J. Angelis	Southern Illinois University
Alison d' Anglejan-Chatillon	University of Montreal
Russell N. Campbell	University of California at Los Angeles
John Haskell	Temple University-Japan

A Survey of Academic Demands Related to Listening Skills

Donald E. Powers

Educational Testing Service

Princeton, New Jersey

RR-85-48

Copyright (c) 1985 by Educational Testing Service. All rights reserved.

Unauthorized reproduction in whole or in part is prohibited.

TOEFL and ETS are trademarks of Educational Testing Service, registered in the U.S.A. and in many other countries.

Abstract

A literature review was conducted in order to identify various parameters underlying listening comprehension. The results of this review were used as a basis for a survey of faculty in six graduate fields as well as undergraduate English faculty. The purpose of the survey was to (a) obtain faculty perceptions of the importance to academic success of various listening skills and activities, (b) assess the degree to which both native and nonnative speakers experience difficulties with these skills or activities, and (c) determine faculty views of alternative means of evaluating these skills.

Faculty perceived some listening skills as more important than others for academic success. These included nine skills in particular that were related primarily to various aspects of lecture content (e.g., identifying major ideas and relationships among them). As might be expected, faculty perceived that nonnative students experience more difficulty than native students with all listening activities, and that nonnative students have disproportionately greater difficulty with some activities, such as following lectures given at different speeds and comprehending or deducing the meaning of important vocabulary. With respect to measuring listening comprehension, some general approaches and specific item types were judged to be more appropriate than others. These included tasks that entail answering questions involving the recall of details as well as those involving inference or deductions.

The results of the survey are used to suggest further research on the construct validity of the Listening Comprehension section of the TOEFL® test.

Table of Contents

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	xi
Previous Research	1
METHODS	3
Questionnaire Development	3
Questionnaire Content	4
Sample Selection	4
Data Collection	5
Data Analysis	5
RESULTS	5
Sample Description	5
Importance of Four Language Skills	6
Importance of Specific Listening Skills	6
Problems Encountered by Native and Nonnative Speakers	7
Differences between the Problems of Native and Nonnative Speakers	8
Description of Typical Courses	8
Appropriateness of General Tasks for Evaluating Listening Skills	9
Appropriateness of Specific Tasks for Evaluating Listening Skills	10
Appropriateness of Various Stimuli for Measuring Listening Comprehension	11
Appropriateness of Various "Speech Events"	11
Critical Incidents	12
SUMMARY AND DISCUSSION	12
Limitations	15
Possible Further Research	16
REFERENCES	19
FOOTNOTES	22
TABLES	23
FIGURES	42
APPENDIX	

List of Tables

		Page
Table 1.	Summary of Selected Research on Variables Related to Listening Comprehension and Their Relevance to the Current TOEFL	23
Table 2.	Ratings of Importance for Academic Success of Listening, Reading, Speaking, and Writing Skills by Discipline	26
Table 3.	Ratings of Importance of Specific Listening Activities by Disciplines	27
Table 4.	Ratings of the Degree to Which Native Speaking and Nonnative Speaking Students Encounter Problems with Specific Listening Activities	29
Table 5.	Descriptions of Typical Courses	33
Table 6.	Ratings of Appropriateness of Various General Tasks for Evaluating Listening Skills of International Students	35
Table 7.	Ratings of Appropriateness of Various Specific Tasks for Evaluating the Listening Skills of International Students	37
Table 8.	Ratings of Appropriateness of Various Stimuli for Measuring Listening Comprehension of International Students	40
Table 9.	Ratings of Appropriateness of Various "Speech Events" for Evaluating the Listening Skills of International Students	41

List of Figures

		Page
Figure 1.	Ratings of Importance of Listening, Reading, Speaking, and Writing Skills	42
Figure 2.	Ratings of Listening Activities: Importance and Problems for Nonnative Speakers	43
Figure 3.	Ratings of Listening Activities: Problems of Native vs. Nonnative Speakers	45
Figure 4.	Possible Configuration for Data Collection	47

Acknowledgments

Numerous people contributed in a variety of important ways to this study. We thank:

- the members of the 1983 TOEFL Research Committee and its Chair, G. Richard Tucker, who gave valuable suggestions regarding the focus of the study
- the study coordinators at each of the 28 participating institutions, and the responding faculty members at these institutions, who provided our data
- project advisors, Joan Morley and Nancy Works, who shared their considerable expertise, particularly in the design of the survey instrument
- Charles Stansfield and Barbara Suomi, who provided helpful reviews of our questionnaire
- Frieda Hardy, who helped to prepare and process the data, Inge Stiebritz, who analyzed the data, and Lorraine Simon, who oversaw numerous administrative aspects of the project including the preparation of the final report
- Sybil Carlson, who shared the valuable experiences gained in a previous study of writing skills, and Gordon Hale, the TOEFL Research Coordinator
- Robert Boldt, Vera Jones, Charles Stansfield, Barbara Suomi, and Kenneth Wilson, who provided constructive reviews of an earlier draft of this report and
- the members of the 1985 TOEFL Research Committee, chaired by Henry Holtzclaw, who made numerous additional suggestions for improving the final report.

The Listening Comprehension section of the TOEFL® test is described (Educational Testing Service, 1981) as measuring the ". . . ability to understand English as it is spoken in the United States." (p. 7) The test items emphasize vocabulary that is frequent in spoken English, structures that tend to be found only in spoken English, and sound/intonation distinctions that nonnative speakers find difficult. Three specific kinds of comprehension questions are employed in the listening section, as follows:

1. Questions for which the examinee must choose the printed statement that is closest in meaning to a recorded statement. (The oral statement is spoken only once.)

2. Questions based on short conversations between two speakers. In response to a question (which is spoken only once) about the conversation, the examinee must choose the most appropriate of four written responses.

3. Questions based on short talks or conversations based on speech that students in the U.S. or Canada might encounter in an academic situation. The examinee must choose the correct response from four printed choices.

The skills measured by these three kinds of questions are not only logically distinct from the item types in the other sections (vocabulary/reading and structure/written expression) of the TOEFL test, but empirically distinguishable as well (Swinton & Powers, 1980).

The major goal of this study was to work towards a more comprehensive validation of the Listening Comprehension portion of the TOEFL test. This was to be aided by the identification of a taxonomy of listening skills, which would provide the substance for a survey of faculty involved in teaching nonnative students. The purpose of the survey was to gather faculty perceptions of (a) the importance to academic success of various listening skills or activities, (b) the degree to which both native and nonnative speakers encountered difficulties with them, and (c) the appropriateness of various means of assessing these skills. Bridgeman and Carlson (1983) have provided this information for writing skills. Angelis (1982) also provided information related to the tasks faced by nonnative students, but considered general deficiencies related to several areas of proficiency rather than any one particular area, such as listening skills.

The information gathered by the survey was intended to provide the basis for the possible development of criterion tasks against which further validation of the current TOEFL Listening Comprehension section could be undertaken. The results might also provide a basis for any future modifications to the TOEFL listening section, if any such modifications were to be contemplated, as well as a context for the continuing evolution of TOEFL test specifications.

Previous Research

In 1971 John Carroll (1971) conducted a very comprehensive literature review on the comprehension of meaningful verbal discourse. He noted an

apparent lack of any comprehensive theory of listening behavior, as well as of any studies (apart from research on speech intelligibility) that had delineated the parameters of listening behavior. Devine (1978) concluded his summary of fifty years of research on listening in much the same vein by noting the lack of any simple, completely acceptable definition of listening. This lack of agreement seems to extend to virtually every aspect of the listening process, including the dimensions that underlie it (Larson, Backlund, Redmond, & Barbour, 1978). Furthermore, as Richards (1983) and Works (personal communication) have noted, there is very little direct research on listening comprehension in a second language: most of what is known derives from research on native listeners.

Despite a general lack of conceptual clarity, some strides have been made towards defining particular aspects of listening comprehension and identifying underlying parameters that need to be considered in any attempts to assess listening skills. Table 1 provides a brief summary of some of the factors thought to affect listening comprehension. These factors relate to the stimulus material, to the speaker, and to the context or other situational variables. Stimulus-related factors include vocabulary load, mode of presentation (e.g., whether or not visual cues are provided), abstractness of material, "vagueness," and degree to which speech is extended. Speaker-related variables include the manner of delivery, the style/skill of the speaker, accent/dialect, rate of speech, and the degree of pauses, redundancies, and slurring. Other contextual variables include such factors as the attentional disposition of the listener, whether or not there are competing stimuli, and whether or not note-taking is allowed. Although there is generally no consensus on the degree to which each of these factors is implicated in listening comprehension, there does seem to be enough evidence to suggest the appropriateness of considering these factors in both teaching and assessing listening skills.

Various efforts have been made towards establishing taxonomies of the roles, functions, and settings of communication skills. Efforts have been made, for example, by Allen and Brown (1976), Bassett, Whittington, and Staton-Spicer (1978), Levison (1976), Richards (1983), and Wiemann and Backlund (1980). Of these, Richards' (1983) effort is the most recent, the most relevant to listening in particular, and the most closely related to the problems of nonnative speakers of English. Richards (1983) has set forth a taxonomy of microskills involved in both conversational and academic listening (or listening to lectures). The academic microskills have provided the basis for Question III of our "TOEFL Survey of Academic Listening Skills" questionnaire (See Appendix).

Other useful guidelines include such schemes as Tinkler's (1980) list of subskills thought to make up listening ability: phoneme discrimination, word recognition, identification of stress, intonation and syntactic patterns, retention of information, anticipation of conclusions, evaluation of ongoing speech, and assessment of the communicative intentions of the speaker. Taylor (1981) has specified types of language to which students are exposed, ranging from cryptic comments to formal lectures. Richards (1983) has given examples of four types of "speech events": conversation, lecture, discussion, and debate. Morley (1972) has defined aural

comprehension as encompassing not only basic auditory discrimination and aural grammar, but also re-auditorizing, extracting vital information, remembering it, and relating it. She has also identified several concept areas to which academic listening skills may be applied. These include numbers/numerical relationships, letters, spelling, abbreviation, sound-letter correspondences, spatial relationships and directions, time and temporal sequences, dates and chronological order, measurements and amounts, and proportions and comparisons.

In summary, there have been a number of efforts to define and research various aspects of the listening process. Although there seems to be no universally agreed-on concept of listening behavior that can completely guide the construction of listening measures, there does appear to be a sufficient basis on which to plan further empirical study of listening skills.

METHODS

The study was modeled closely after a survey of the academic writing tasks that are required of graduate and undergraduate foreign students (Bridgeman & Carlson, 1983). That study provided a model not only because the methods employed were effective but also because the use of generally similar procedures would provide information for listening skills that would be somewhat comparable to that obtained for writing abilities.

Questionnaire Development

A literature review was undertaken to identify specific areas to be included on the questionnaire. The review was aided by computer searches, conducted by the ERIC Clearinghouse on Languages and Linguistics, of the literature on teaching listening comprehension. Also, the ERIC Clearinghouse on Tests, Measurement, and Evaluation was used to locate existing measures of listening comprehension and other aural language skills. In addition, two project consultants (Joan Morley at the University of Michigan and Nancy Works at the University of Illinois) provided valuable clues to other relevant work, including their own reseach on the teaching and testing of listening comprehension (e.g., Works, 1982 and Morley, 1972; 1976).

One highly relevant article (Richards, 1983), published shortly after the project was initiated, proved to be extremely valuable in developing the questionnaire. Richards (1983) identified 33 skills required for conversational listening and 18 relevant to academic listening (listening to lectures). This list of academic skills was used extensively in questionnaire design. In addition, Richards (1983) also identified eight generic exercise types that have been used to assess listening skills. This list was included almost verbatim in the questionnaire.

A number of specific kinds of listening comprehension questions were also identified through various sources (e.g., Aitken, 1979; Otto, 1979), including the TOEFL test itself, and examples of each kind were given in the questionnaire. For some question types it was possible to include an actual example. For others, however, the length of questions made it necessary simply to describe the question type rather than present a complete example. In addition, in order to simplify data collection we had

questionnaire respondents read the questions rather than listen to them. Therefore, as suggested below, these methods of presentation may have affected (in unknown ways) the ratings of the various question types.

Several other resources proved to be extremely useful in characterizing various aspects of academic lectures. These included Mason's (1983) "Understanding Academic Lectures" and Brown, Bakhtar, and Youngman's (1984) recent work on developing a typology of lecturing styles.

Questionnaire Content[1]

The questionnaire included eight major sections designed to elicit information on:

I. Numbers and geographic background of students.

II. Importance of listening, reading, speaking, and writing skills.

III. Importance of various listening activities to academic success, and the degree to which both native and nonnative speakers encounter problems with these activities.

IV. Features of typical courses, especially the characteristics of lectures.

V. Appropriateness of various general kinds of tasks for evaluating the listening skills of international students.

VI. Appropriateness of various specific kinds of questions for evaluating the listening skills of international students.

VII. Appropriateness of various stimuli (persons/situations) for testing listening comprehension.

VIII. Appropriateness of various "speech events" for testing listening skills.

Faculty were also asked to list briefly any "critical incidents" involving nonnative students' understanding (or misunderstanding) of material presented orally.

Sample Selection

The same 34 institutions that participated in the Bridgeman and Carlson (1983) survey constituted the basic sample. Briefly, these institutions had been selected because they enrolled large numbers (more than 1,000) of nonnative students. About equal numbers were located in each of the four U.S. geographic regions (Pacific, South/Southwest, Northeast, and Midwest) having the highest proportions of nonnative students. Three Canadian institutions with high proportions of nonnative students were also included.

The same academic disciplines as those selected by Bridgeman and Carlson (1983) were also chosen for this study. At the graduate level, six disciplines--electrical engineering, civil engineering, psychology, chemistry, computer science, and business--were selected because each of

these areas tends to enroll relatively large numbers of nonnative students and because they represent quite different areas of graduate education. Bridgeman and Carlson (1983) included undergraduate English departments because most writing by undergraduates is done in these departments. This area was included in the current study primarily to maintain comparability with Bridgeman and Carlson's (1983) results, even though its inclusion in a study of listening skills may be somewhat less appropriate than in one of writing skills.

Data Collection

Each of the individuals who served as a data collection coordinator for Bridgeman and Carlson was recontacted to ascertain his or her willingness to participate in a similar study of listening skills. In some cases, the requests were made of individuals who had replaced the initial coordinators. Again, these coordinators were asked to identify one faculty member in each of the seven disciplines who would be familiar with the academic demands of nonnative students, and to distribute, collect, and return the questionnaires. Each coordinator and faculty member was paid a small honorarium for participating in the study.

Data Analysis

Results for each question were subjected to a one-way analysis of variance with major field as the single factor. No multiple comparisons among means were attempted, and significant differences among fields were attributed to differences from particular fields only when such differences were clearly responsible for overall significance.

RESULTS

Sample Description

Of the 34 institutions contacted initially, a total of 28 agreed to participate in the study. A total of 150 questionnaires were returned from these schools. All but four questionnaires from one institution were returned in time to be included in the analyses. Responses were received from graduate faculty in civil and electrical engineering (N = 37), psychology (N = 21), chemistry (N = 23), computer science (N = 21), and business (N = 22), and from 22 undergraduate faculty in English. (Note that in all subsequent analyses civil and electrical engineering have been combined because preliminary review suggested the similarity of responses in these two related disciplines.)

All faculty reported that their departments enrolled at least some nonnative students. The median numbers of nonnative students reported for graduate departments ranged from 17 for psychology and 29 for chemistry to 77 for computer science and 85 for engineering; the median nonnative student enrollment for undergraduate English departments was 120 students. (It should be noted that undergraduate faculty were asked to give these numbers for all first year nonnative students at their institutions, not just English majors.)

Because of the different numbers of students enrolled in the various disciplines, the median reported percentages of nonnative students differed significantly by field, ranging from a low of 8 percent for psychology to a high of 34 percent for engineering. The computer science (32 percent), chemistry (31 percent) and business (24 percent) departments in our sample also reported significant percentages of nonnative students. The median reported for undergraduate English departments was 10 percent.

A majority or near majority of respondents in each field (47 percent of psychology faculty to 85 percent of computer science faculty) indicated that most of their nonnative students came from Asian countries. The next most frequently mentioned geographic area from which most students came was the Middle East (16 percent), followed by Latin America (12 percent). These overall figures varied somewhat by discipline, with, for example, Latin America as the second most often mentioned geographic area for psychology (32 percent).

Importance of Four Language Skills

When asked to indicate the importance of listening, reading, speaking and writing abilities for success in their departments, faculty generally gave the highest ratings to reading ability (see Table 2 and Figure 1), which was seen as the most important of the four abilities in all disciplines but English, where it was rated as a very close second to writing. Speaking ability was rated as the least important in all disciplines, although speaking ability was rated higher in business than in other disciplines. Listening was rated as either the second or third most important ability in each discipline, as was writing ability. Writing ability was viewed as the most important ability by English department faculty.

Although the rank order of listening ability fluctuated between second and third across disciplines, the mean ratings did not differ significantly across disciplines ($p > .05$), indicating some degree of consensus regarding the importance of listening skills.

Importance of Specific Listening Skills

Table 3 shows the ratings of importance for 21 specific lecture-related listening activities as they relate to successful academic performance. Nine activities were rated as very important. Briefly, these were:

1. Identifying major themes or ideas
2. Identifying relationships among major ideas
3. Identifying the topic of a lecture
4. Retaining information through note-taking
5. Retrieving information from notes
6. Inferring relationships between information
7. Comprehending key vocabulary

8. Following the spoken mode of lectures

9. Identifying supporting ideas and examples

There was a steady decrease in the mean ratings of importance after these nine activities, although all but five of the remaining activities were also rated as being, on average, at least moderately important (mean = 3.0 or higher). Generally, the lowest ratings were made for such secondary skills as recognizing (a) irrelevant matter[2] (such as jokes), (b) the function of nonverbal signals, (c) attitudinal signals, and (d) functions (e.g., change of pitch) that signal information structure.

There were some differences across disciplines in ratings of importance. For English departments, identifying the role of discourse cues, deducing the meaning of words from context, recognizing the function of such factors as change of pitch, and recognizing attitudinal signals were rated as significantly more important ($p < .05$ or less) by faculty from English departments than from others. Business faculty rated some activities (identifying the role of discourse cues, retaining information through note taking, and retrieving information from notes) as significantly less important than did other faculty.

Problems Encountered by Native and Nonnative Speakers

Table 4 contains ratings (means and standard deviations) of the extent to which faculty perceived both native and nonnative speakers as encountering problems with various listening activities. The first point to note is that, with one exception, average ratings did not differ significantly across disciplines, either for native or for nonnative speakers. This suggested that we could focus on the averages for the total group. As might be expected, nonnative speakers were perceived as having more frequent or more severe problems with each activity than were native speakers.

Figure 2 shows the average ratings of importance across all disciplines for the various listening activities. The activities have been ordered by importance rating. Also displayed are the ratings of the degree to which nonnative speakers encountered problems with each activity. With the exception of "identifying the topic of a lecture", each of the nine most important activities was also rated at least moderately high (greater than 3.0) with respect to the degree to which nonnative speakers experienced difficulty.

Of the 12 listening activities that were seen as less important to academic success, five were perceived as activities with which nonnative speakers encountered at least moderate (greater than 3.0) difficulty. Thus, although apparently less important, these activities may still be appropriate targets of assessment because of the difficulty experienced by nonnatives. These were:

1. Deducing the meanings of words from context

2. Understanding informally structured lectures

3. Following lectures given with different

   a. speeds and
   b. accents or dialects and

4. Recognizing irrelevant matter, such as jokes or digressions.

Differences between the Problems of Native and Nonnative Speakers

Figure 3 shows the degree to which native and nonnative speakers were seen as having problems with each of the listening activities. Activities have been ordered by degree of problem for native speakers. As can be seen, faculty did perceive some differences among the activities with respect to the degree to which speakers encountered problems, although this distinction was less clear than was the degree to which these activities were thought to be differentially important. As mentioned before, nonnative speakers were perceived as encountering greater difficulty than native speakers with each activity. They also were seen as having disproportionately greater difficulty with the following activities:

1. Following lectures given with different speeds

2. Comprehending key vocabulary related to the topic of a lecture

3. Deducing the meaning of words from the context in which they appear

4. Understanding informally structured lectures

5. Identifying the role of discourse cues

6. Following different modes of spoken lectures

7. Following different modes of audio lectures

8. Recognizing irrelevant matter

9. Recognizing the function of such features as change of pitch and sentence stress

Description of Typical Courses

Because the various features of typical courses might have some relevance for determining appropriate targets of listening assessment, faculty were asked to indicate how well certain features described the courses they typically taught. Table 5 contains these ratings.

Overall, each of three possible purposes of lectures was characterized as being at least somewhat descriptive of typical lectures, suggesting that classroom lectures often serve several purposes. English and chemistry faculty were slightly more likely than psychology faculty to use lectures to interpret material from an assigned text or reading. There were no other significant differences across disciplines in the purposes of lectures.

With respect to the roles that lecturers played, English and business faculty were significantly ($p < .001$) more likely to lecture in an interactive mode--to facilitate discussion--than were chemistry and computer science faculty, who more typically employed a noninteractive mode--to present information with a minimum of exchange.

English faculty were significantly more likely ($p < .01$) to report the frequent use of examples or anecdotes than were their colleagues in other disciplines, and chemistry faculty were significantly ($p < .01$) more likely to use a deductive style of presentation. Finally, English faculty were significantly ($p < .01$) more likely to make frequent use of handouts than were psychology faculty.

Appropriateness of General Tasks for Evaluating Listening Skills

Table 6 contains ratings of several general sorts of tasks that have been (or could be) used to assess listening skills. As can be seen, faculty were able to discriminate among these tasks, with average ratings for all faculty ranging from 3.12 to 4.35. While no single task was seen as inappropriate (all average ratings were greater than 3.0), answering questions involving recall of details and those involving inferences and deductions were viewed as somewhat more appropriate than were the other tasks mentioned, as was condensing, i.e., being able to reduce what is heard to an outline of main points. The task of "matching or distinguishing," i.e., choosing a picture to correspond with what was heard, was seen as least appropriate. (The same task using a written response, however, was rated as somewhat more appropriate.) "Scanning," i.e., listening for selected specific pieces of information, and "extending," i.e., going beyond what was said, were also seen as somewhat less appropriate.

There were some differences across disciplines in the perceptions of these general tasks. Using pictorial responses in the matching/distinguishing task was significantly different ($p < .01$) across fields, being rated highest by chemistry faculty and lowest by business faculty. Certain other tasks (transcribing, skimming, condensing, and answering questions involving the recall of details) also differed significantly ($p < .05$) across disciplines.

Faculty sometimes volunteered reasons for their perceptions of certain tasks as inappropriate. The matching/distinguishing task (especially pictorial) was viewed as "too low level," "lacking content," "never encountered," and "too elementary." Pictorial responses were thought to be inappropriate for assessment based on financial or psychological content, and difficult to devise for engineering concepts. The same feeling was expressed in English departments, which concentrate on abstract concepts.

One computer science faculty member felt that, in his field, symbols could be used more appropriately than pictures. Others expressed the feeling that choosing a pictorial response was not a verbal exercise.

"Scanning" was also thought of generally as a lower-level comprehension skill. Faculty expressed concern that course material in their disciplines needs to be considered "in detail and in depth" and that scanning results in ignoring important details. Scanning of audio material was viewed as an infrequent activity, and, as one computer science faculty member commented, course material is too dense for mere scanning. The same basic concern was expressed regarding the technical details of engineering lectures. Another concern was that scanning may imply that students know beforehand what relevant information is to be presented in lectures, and that this is simply not often the case.

"Extending or constructing," i.e., going beyond what was said, was often viewed as inappropriate because it requires abilities beyond or other than listening skills. Also, as one chemistry professor noted, at the beginning graduate level, grasping concepts as presented and remembering factual information are more important than being able to go beyond what is given.

There were also, however, minority views expressed regarding the tasks rated as most appropriate. For example, "answering" was thought to depend more on memory than on comprehension. As one engineering faculty member noted, "Nonnative students are good at copying down and repeating, but poor at extending and applying." One computer science professor noted that "the details are in the books," and several others remarked about the greater appropriateness of concepts over details.

There were extremely few specific comments regarding why "condensing" might be inappropriate. Besides "matching/distinguishing," "scanning," and "extending/constructing," the most frequently expressed concerns were directed at "predicting," i.e., guessing or predicting outcomes, and "transcribing," e.g., dictation. Several comments suggested that "predicting" was also seen as involving skills beyond listening and possibly other background information as well.

"Transcribing" was generally seen as a low-level skill--that it does not adequately reflect abilities required for academic success or that success on this task does not reflect a student's understanding. As one English faculty member commented, "This seems more like training for secretaries."

Appropriateness of Specific Tasks for Evaluating Listening Skills

Again, each of several specific kinds of tasks or question types (see Table 7) was rated as at least moderately appropriate for measuring listening skills, and, with one exception, there were no significant differences across disciplines in faculty perceptions of the appropriateness of these tasks. (The appropriateness of pictorially presented comprehension questions differed significantly [$p < .001$], with chemistry and computer science faculty rating these as more appropriate than did English and business faculty.)

"Number questions," i.e., questions involving the comprehension of numbers and numerical concepts, received the highest overall rating of appropriateness. These high ratings were especially prominent in the quantitative fields of chemistry and engineering, but appeared even in the less quantitative fields. "Paraphrase answer" questions, i.e., selecting the best paraphrase of a statement, "appropriate response" questions, i.e., selecting the most logical answer from several available choices, and "conversation questions (or dialogues)," i.e., comprehending a conversation between speakers, were the question types that were rated as the next most appropriate. (We note that the example given for "number questions" was also a "paraphrase answer" question. Thus, the rating for "number questions" may also reflect the confounding of numerical content with question format.)

Each of two variations of "dictation" questions and each of two variations of "aural cloze" questions, especially the nonmultiple choice fill-in, were rated as the least appropriate of any question types. The most frequently cited reason for the inappropriateness of dictation exercises was that students do not typically need to copy lectures verbatim, but rather they need to comprehend and organize material. Dictation was seen as something that simply is not ordinarily required of students.

Aural cloze questions were viewed as inappropriate because a listening cloze test would be too confusing and difficult for foreign students. Potential problems with scoring on a large scale and the likely difficulty of correctly keying answers were mentioned also. In addition, several faculty believed that cloze tests entailed more production skills (writing) than receptive (listening) skills.

Appropriateness of Various Stimuli for Measuring Listening Comprehension

Table 8 shows that faculty in each of the six disciplines generally regarded each of several possible stimuli as relatively appropriate for assessing the listening comprehension of nonnative students. The relatively low ratings by business faculty of "a librarian," "a laboratory or teaching assistant," and "a student health service nurse or doctor" as stimuli accounted for the only significant differences among disciplines. "A professor giving a classroom lecture" was rated as very appropriate by nearly every rater, and was chosen as the most appropriate of all stimuli by a significant majority of faculty in each discipline. "A fellow student," "a student health service worker," and "a registrar" were viewed as least appropriate, but, as mentioned, even these were regarded as quite appropriate.

Appropriateness of Various "Speech Events"

With the exception of "audiovisual presentations," each of the various "speech events" mentioned in Table 9 was rated as relatively appropriate as a basis for the assessment of listening skills. "Classroom lectures," "discussions," and "laboratory instructions" received slightly higher ratings than did others. Only for "laboratory instructions" did ratings differ significantly across fields, with low ratings from business faculty accounting for this difference.

Critical Incidents

When asked to describe briefly "critical incidents" involving nonnative students' understanding (or misunderstanding) of lectures, and so on, faculty in each discipline most often mentioned the misinterpretation of assignments or other instructions such as those given in laboratories or those involved in institutional procedures (e.g., how to submit a thesis). Also frequently mentioned were difficulties in taking examinations, e.g., failing to interpret questions properly, or to follow instructions for guessing or for pacing oneself on an exam.

Faculty in each discipline also often mentioned problems in understanding idioms or American slang and technical terms such as those used in computer science. Nonnative students were noted to tend to interpret nonliteral language literally and to misinterpret, for example, irony and humorous remarks.

Faculty said that they based their evaluations of nonnative students' listening skills on such factors as:

1. Students' failure to ask questions in class

2. Students' performance on course examinations

3. Students' contact with instructors after classes

The failure to ask questions in class was interpreted by some faculty as a lack of understanding. (Other faculty, however, saw this as more of a cultural phenomenon than a lack of comprehension.) Faculty quite often mentioned heavy contact with nonnative students after class hours as an indication of the difficulties students were encountering, reporting, for example, that nonnative students sometimes "monopolized" scheduled faculty office hours in attempts to obtain clarification of class lecture material. Numerous incidents were also reported that involved students returning after class with English-speaking compatriots as interpreters.

Faculty also alluded to problems that nonnatives had with the pace of lectures. One faculty member cited professors who speak too rapidly as a problem for nonnatives, and one commented on a nonnative student who tried consistently to review unfamiliar terms during lectures. Faculty also mentioned a variety of other specific problems or incidents.

SUMMARY AND DISCUSSION

- Of the four broad language skills, reading was consistently rated as the most important for academic success in each of the six disciplines considered in the study. Ratings of listening were also very consistently high across disciplines, and usually lower only than those for reading. Writing ability was viewed as very slightly less important than listening overall, and speaking skills were generally rated as the least important of the four major language skills. In contrast to reading and listening,

which were rated as consistently high across disciplines, perceptions of the importance of writing and speaking varied significantly from discipline to discipline. This finding, i.e., the perception of reading (especially) and listening skills as somewhat more important than writing and speaking skills (especially), is generally consistent with the findings of other surveys of less representative samples of faculty (e.g., Johns, 1981).

What this implies is simply that reading, listening, and writing skills should probably continue to be important targets of assessment in the TOEFL program. Because the importance of writing and speaking skills appears to be somewhat less consistent across disciplines, providing separate measures of these abilities, as is now accomplished by the Test of Spoken English, for example, would seem appropriate. This strategy might also be appropriate for the planned expansion in 1986 of the TOEFL program to include a direct measure of writing.

- Faculty willingly offered their opinions of the importance of various listening skills and activities, and they were able to distinguish among various skills/activities with respect to their importance for academic success. Faculty perceived nine listening skills in particular as somewhat more important than others. These skills, some of which are already represented in the current test, would seem to be appropriate targets of assessment for the TOEFL program, and their importance might be emphasized further as the specifications for the TOEFL listening section continue to evolve.

- Relatively few differences of opinion were noted among faculty in different disciplines with respect to the importance of specific listening skills. Faculty in English, however, generally placed somewhat more importance on the more subtle aspects of listening comprehension, such as recognizing the role of discourse cues, than did faculty in other disciplines. This finding would seem to suggest that considerable latitude could be exercised in selecting specific skills to be assessed in the listening section of TOEFL, while still maintaining it as appropriate for a variety of disciplines.

- As expected (and as judged from several faculty comments), the task of rating the problems encountered by nonnative students was a difficult one. Nonetheless, most faculty volunteered their opinions. The patterns of ratings suggested, however, that faculty may indeed have had difficulty in determining when, and with which activities, students experienced difficulties. Specifically, faculty members differentiated less among the various listening skills and activities on this dimension, i.e., the degree of difficulty encountered by students, than they did with respect to the importance of these skills.

- Nonnative students were seen consistently as having greater difficulty with each of the various listening activities than were native speakers. In addition, some activities were judged to be <u>disproportionately</u> more difficult for nonnatives than for native speakers in relation to other activities. There were, however, virtually no differences among disciplines with respect to faculty perceptions of the degree to which

either native or nonnative students experienced difficulties with each of the various language activities. Again, this finding is encouraging with regard to the need to offer measures, like TOEFL, that must serve multiple constituencies.

- With few exceptions, the listening activities perceived as most important tended also to be those with which nonnative students were thought to have the greatest difficulty. This finding is also encouraging with regard to identifying the most appropriate targets of assessment.

- While none of the general tasks or specific item types listed was judged as generally very inappropriate for assessing the listening skills of nonnative students, faculty did make relatively crisp discriminations among the various tasks and question types, and judged some to be significantly more appropriate than others. This finding suggests that, all other things being equal, some kinds of tasks and question types would be preferred over others.

- As might be expected, lectures were perceived by faculty in all disciplines as being by far the most appropriate stimulus for measuring listening comprehension. Lectured-based questions would seem, therefore, to be a good basis for any new TOEFL test question types or for criterion tasks against which the current TOEFL could be gauged. However, because the features of lectures and the role and style of lecturers vary from discipline to discipline according to faculty, care should be exercised in balancing lecture-based question types so as to represent fairly the kinds of lectures encountered in the various disciplines. In particular, the degree of student-teacher interaction and the use of examples or anecdotes, as well as the predominant style of presentation (deductive or inductive), should be carefully considered.

- There were relatively few differences across disciplines in the degree to which the various general tasks and specific listening question types were judged as appropriate. However, whether or not response options were presented pictorially was a major differentiating factor, with faculty in the more quantitative/scientific disciplines like chemistry and computer science rating pictorial response questions as more appropriate than did their counterparts in other disciplines.

- Reasons offered by faculty for the relative inappropriateness of some tasks related largely to (a) task difficulty, (b) the extent to which tasks correspond with what is required of students, and (c) the extent to which tasks appear to measure listening skills instead of other language or nonlanguage skills. From the standpoint of general acceptance of such tests as TOEFL, these would seem to be important factors to continue to consider in test development.

- The three kinds of listening comprehension questions used currently in the TOEFL were rated as being among the most appropriate of those considered[3], as were questions involving the comprehension of numbers and numerical concepts. The endorsement of currently employed TOEFL question types is certainly encouraging. The very high rating of numerical

questions was somewhat unexpected, especially since, until recently, questions involving numbers and numerical concepts were avoided on TOEFL because such questions seemed to test matters other than language. Currently, numerical questions are used, if this information is repeated. It is useful to know that the more recent use of such questions is not perceived negatively by faculty, and, therefore, that such questions might continue to play some role in future versions of TOEFL.

Limitations

This study appears to have generated some useful information, and our efforts "to improve the understanding of foreign students' listening abilities" were "applauded" by several respondents. However, when applying the results, certain inherent limitations of the study should be heeded. The notion of analyzing listening activities may have been "foreign" to many faculty members who were not involved intensely in language instruction or testing. In particular, such concepts as "discourse cues" and "nonverbal signals" may be somewhat far afield for faculty in non-language disciplines. Moreover, while the rating of such passive, nonobservable skills as listening may be difficult generally, non-language-oriented faculty may have even greater difficulty in determining when students encounter specific kinds of problems. Finally, although we did get some information regarding the bases on which faculty judged students' listening skills, we were not able to quantify the nature and amount of student/faculty interaction--a variable that is likely to moderate such judgments.

One consultant (Joan Morley) suggested possible alternative data-collection procedures that might help overcome non-language-oriented faculty members' lack of familiarity with linguistic/grammatical terms or concepts. These include (a) providing on-site training of faculty respondents and (b) conducting on-site interviews. Other questionnaire respondents suggested that "the real experts"--international students themselves--might provide valuable insights into the problems they encounter, and would therefore be an appropriate focus of any further data collection.

These suggestions perhaps highlight the fact that the data gathered in this study, i.e., faculty perceptions, constitute only one of many sources of information that must be considered in developing and validating the various sections of TOEFL. Comparable information from international students or from ESL specialists might have produced different results. We speculate, for instance, that faculty and ESL specialists might use different bases to judge the suitability of various alternative question types. If, for example, ESL experts were to rely more on linguistic considerations and less on appearance and familiarity than do faculty members, then the low ratings given to such tasks as dictation (transcribing) and aural cloze might have been higher.

Possible Further Research

Previous studies, e.g., Swinton and Powers (1980), have demonstrated that the test questions that make up the Listening Comprehension section of TOEFL determine a dimension of performance that is distinct from other dimensions determined by other TOEFL test question types. Although this information is indeed important and reassuring, it does not exhaust the kinds of information that are relevant to establishing the construct validity of the measure. Swinton and Powers' (1980) factor analytic study dealt primarily with the internal structure of the TOEFL, showing that the Listening Comprehension questions comprising the TOEFL relate more closely to one another generally than to questions in other sections of the test. A logical extension of this research would be to relate the TOEFL questions from various sections of the test to other, external measures of the skills these items are thought to reflect. These skills should be measured, if feasible, by other methods as well as by other measures.

Bachman and Palmer (1983) have reported a construct validity study of the FSI Oral Interview that Oller (1983) has judged to be "...one of the most significant advances in language testing research over the last several decades." (p. 154). The most significant aspects of this research were thought to be the use of a multitrait-multimethod approach and the application of confirmatory factor analytic methods. The multitrait-multimethod approach enables the examination of both convergent validity, i.e., the extent to which different measures of the same trait tend to agree, and discriminant validity, i.e., the extent to which the measurement of different traits by the same method tend not to agree. The use of confirmatory factor analysis enables an examination of how well each of several alternative hypothetical models can explain the relationships among the various measures.

Frederiksen (1982) has suggested a procedure for test development and validation that begins with a "scientific investigation of the performance to be assessed or predicted" (p. 2), or a task analysis based on any of several methods, including observing the criterion performance that is to be predicted. The next suggested step is to develop a criterion test, for which cost and efficiency are not major concerns, since this test is not necessarily to be used on a large scale. The primary concern is that the test elicit behaviors that are as similar as possible to those observed in the real-life criterion situation. Subsequent steps involve the development and evaluation of an assessment test that is based on observations about criterion performance.

Instrument development. It is recommended that the spirit of Frederiksen's approach be followed based on the results of the survey reported here, which has provided task-related information. The selection or development of criterion measures should be based on such survey findings as:

(a) the uniformly high ratings given to lectures as stimuli for assessing listening comprehension,

(b) the rating of some listening activities as being more critical for academic success, as well as more difficult for nonnative students, and

(c) the assessment of some general tasks, particularly condensing and note-taking, and some specific question types as being more appropriate than others.

(Such sources as Annand's (1979) compendium of five-minute lectures and associated questions for listening comprehension might be a useful resource for identifying lecture materials.)

In addition, although some specific kinds of questions, in particular aural cloze and dictation, were not judged by faculty to be especially appropriate, these question types are not without merit. Dictation, for example, can be an effective test of language proficiency (Aitken, 1979), and the cloze procedure has traditionally been seen as a powerful means of assessing language proficiency. Apparently, relatively little research has focused on aural cloze tests of second language proficiency, but Morley (1976) has developed a modified aural cloze procedure that may deserve consideration. Although, from a face validity perspective, cloze and dictation methods may be less acceptable than others, they may nonetheless be appropriately included as alternative measures of listening skills in construct validation efforts. Such factors as the speed with which material is presented in these exercises could be systematically manipulated.

An additional approach that should be seriously considered, as an alternative method of assessment, is student self-ratings of various facets of their English language proficiency. Again, while self-ratings would probably be of little if any value in an operational testing program, they would provide another means of assessment to be considered in a multitrait-multimethod framework.

One possible configuration of instruments is suggested in Figure 4. Although this configuration would not permit the "definitive" construct validation of all the language proficiencies considered in the TOEFL program, it would extend considerably what is known about the current TOEFL examination.

All new measures would, however, need to be pilot tested.

Data collection. It is suggested that consideration be given to collecting data for alternative measures of listening comprehension on the same day that examinees take the TOEFL test. To the extent possible, the same procedures used by Carlson, Bridgeman, Camp, and Waanders (1984) would be employed. Basically, these entailed gaining the cooperation of TOEFL test center supervisors or agents in several countries. Carlson et al. (1984) collected data in two Chinese-speaking, three Arab-speaking, and three Spanish-speaking countries.

Data analysis. The correlation among the various alternative measures would be examined for evidence that they relate to each other in logically meaningful ways. According to the standard criteria for multimethod-multitrait analyses, judgment would be based on the extent to which different measures of the same trait tend to correlate highly, etc.

For the confirmatory factor analysis, alternative models or factors would be specified and tested to determine the one having the most explanatory power. As an example, one model might specify a single general language proficiency factor; others might include separate factors, either correlated or uncorrelated, for each trait.

Expected results. The results of this work would shed further light on the meaning of the TOEFL measure, particularly the Listening Comprehension section, by revealing relationships with various non-TOEFL measures of language proficiency. In addition, the development and administration of several experimental measures of listening comprehension should provide data that would enable the evaluation of these measures as possible candidates for inclusion in future versions of TOEFL, if any changes were to be contemplated. Finally, information on how TOEFL Listening Comprehension scores relate at various levels to perfomance on other measures of listening comprehension would provide a start to anchoring the current TOEFL Listening Comprehension scale in terms of other measures of similar skills.

References

Aitken, K. G. (1979). Techniques for assessing listening comprehension in second languages. Audio-Visual Language Journal, 17, 175-181.

Allen, R. R., & Brown, K. L. (Eds.). (1976). Development of communication competence in children. Skokie, Illinois: National Textbook Company.

Angelis, P. J. (1982). Language skills in academic study. Princeton, NJ: Educational Testing Service.

Annand, W. H. (1979). Lectures for listening comprehension. Washington, DC: American Language Academy.

Bachman, L. F., & Palmer, A. S. (1983). The construct validity of the FSI oral interview. In J. W. Oller, Jr. (Ed.), Issues in language testing research. Rowley, Massachusetts: Newbury House Publishers, Inc.

Bassett, R. E., Whittington, N., & Staton-Spicer, A. (1978). The basics in speaking and listening for high school graduates: What should be assessed? Communication Education, 27, 293-303.

Beighley, K. C. (1952). Experimental study of the effect of four speech variables on listener comprehension. Speech Monographs, 19, 249-258.

Beighley, K. C. (1954). Experimental study of the effect of three speech variables on listener comprehension. Speech Monographs, 21, 248-253.

Bridgeman, B., & Carlson, S. (1983). Survey of academic writing tasks required of graduate and undergraduate foreign students (TOEFL Research Report No. 15). Princeton, NJ: Educational Testing Service.

Brown, G. A., Bakhtar, M., & Youngman, M. B. (1984). Toward a typology of lecturing styles. British Journal of Educational Psychology, 54, 93-100.

Carlson, S. B., Bridgeman, B., Camp, R., & Waanders, J. (1984). Relationships of admissions test scores to writing performance of native and nonnative speakers of English (Draft final report to the TOEFL Research Committee). Princeton, NJ: Educational Testing Service.

Carroll, J. B. (1971). Learning from verbal discourse in educational media: A review of the literature (ETS Research Bulletin RB-71-61). Princeton, NJ: Educational Testing Service.

Devine, T. G. (1978). Listening: What do we know after fifty years of research and theorizing? Journal of Reading, 21, 296-304.

Educational Testing Service. (1981). TOEFL test and score manual. Princeton, NJ: Educational Testing Service.

Eisenstein, M., & Berkowitz, D. (1981). The effect of phonological variation on adult learner comprehension. Studies in Second Language Acquisition, 4, 75-80.

Flaherty, S. E. (1979). Rate-controlled speech in foreign language education. Foreign Language Annals, 12, 275-280.

Foulke, E., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. Psychological Bulletin, 72, 50-62.

Frederiksen, N. (1982). Construct validity and construct similarity: Methods for use in test development and test validation (ETS Research Report RR-82-45). Princeton, NJ: Educational Testing Service.

Howe, M. J. A. (1970). Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. Journal of Educational Research, 64, 61-63.

Johns, A. M. (1981). Necessary English: A faculty survey. TESOL Quarterly, 15, 51-57.

Kalivoda, T. B. (1981). Developing advanced listening comprehension skill in a foreign language: Problems and possibilities. Hispania, 64, 80-85.

Larson, C., Backlund, P., Redmond, M., & Barbour, A. (1978). Assessing functional communication. Urbana, Illinois: ERIC Clearinghouse on Reading and Communication Skills; and Falls Church, Virginia: Speech Communication Association.

Levison, G. K. (1976). The basic speech communication course: Establishing minimal oral competencies and exemption procedures. Communication Education, 25, 222-230.

Mason, A. (1983). Understanding academic lectures. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Morley, J. (1972). Improving aural comprehension. Ann Arbor, Michigan: University of Michigan Press.

Morley, J. (1976). Listening Dictation. Ann Arbor, Michigan: University of Michigan Press.

Oller, J. W., Jr. (1983). Issues in language testing research. Rowley, Massachusetts: Newbury House Publishers, Inc.

O'Neill, J. J. (1954). Contributions of the visual components of oral symbols to speech comprehension. Journal of Speech and Hearing Disorders, 19, 429-439.

Otto, S. A. (1979). Listening for note-taking in EST. TESOL Quarterly, 13, 319-328.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. TESOL Quarterly, 17, 219-240.

Rubin, R. B. (1982). Assessing speaking and listening competence at the college level: The communication competency assessment instrument. Communication Education, 31, 19-32.

Swinton, S. S., & Powers, D. E. (1980). Factor analysis of TOEFL for several language groups (TOEFL Research Report No. 6). Princeton, NJ: Educational Testing Service.

Taylor, H. M. (1981). Learning to listen to English. TESOL Quarterly, 15, 41-50.

Tinkler, T. (1980). Learning to teach listening comprehension. English Language Teaching Journal, 35, 28-34.

Travers, R. M. W. (1967). Research and theory related to audiovisual information transmission. Revised edition of report, Western Michigan University, Contract No. 3-20-003, U. S. Department of Health, Education and Welfare.

Weiland, A., & Kingsbury, S. J. (1979). Immediate and delayed recall of lecture material as a function of notetaking. Journal of Educational Research, 72, 228-230.

Wiemann, J. M., & Backlund, P. (1980). Current theory and research in communicative competence. Review of Educational Research, 50, 185-199.

Wilcox, G. K. (1978). The effect of accent on listening comprehension: A Singapore study. English Language Teaching Journal, 32, 118-127.

Works, N. M. (1982). A survey of American materials for the teaching of listening comprehension. Unpublished manuscript.

Yates, V. (1979). Listening and note-taking (2nd ed.). New York: McGraw Hill.

Footnotes

[1]The identification of Richard's (1983) work, the effective early "long distance" use of consultants, and the determination that "lecture comprehension" was primary among the academic demands of nonnative students led to some modifications of initial survey plans. Specifically, it was judged that visits to important centers of listening assessment could be profitably replaced by additional use of two consultants, either by telephone or mail. Also, because the focus of the survey was on lecture comprehension, it became less appropriate to survey nonfaculty members. Instead, a question was included on the appropriateness of various extra-classroom situations for measuring listening comprehension.

[2]The use of the term "irrelevant" was perhaps unfortunate here, because in retrospect it may have introduced a bias towards low ratings. As one reviewer has pointed out, many "asides" may in fact be relevant to the teaching process, if not to the subject matter itself.

[3]The possibility was raised that, because they undergo extensive pretesting and revision, the examples of currently used TOEFL question types may have received higher ratings than did examples of other, less refined item types. This possiblity was dismissed as unlikely, however, because, on inspection, the formats of each sample question type included in the questionnaire seemed to be about equally polished. If each of these examples had been selected <u>at random</u> from a larger pool representing each type, the quality of questions might indeed have been lower for nonpretested questions than for TOEFL questions.

Table 1

Summary of Selected Research on Variables Related to Listening Comprehension and Their Relevance to the Current TOEFL

Variable	Source	Summary	Current TOEFL
Stimulus-related			
Vocabulary load	Carroll (1971)	Commonly recognized that vocabulary load is one factor that makes a listening (reading) test difficult.	Difficult lexical items are avoided.
Presentation mode (audio only or audio and visual)	Carroll (1971) Travers (1967)	Listening and reading is advantageous to some, but may interfere at some levels. The evidence for enhancement of comprehension with pictures is inconclusive. Combined audiovisual presentations are often more difficult to comprehend than single-channel presentations because of possible information overload.	No visual stimuli are used.
Abstractness of material	Beighley (1952, 1954)	Easy and difficult speeches differed in their ratio of abstract to concrete material.	Material varies in abstractness; most would probably be considered as relatively concrete.
"Vagueness"	Carroll (1971)	Overuse of such vague words as "very," "pretty," "some," and "maybe" versus the skillful use of more "explaining links" like "therefore" and "because" may affect comprehension.	Relatively little use of vague words.
Type of interactional event	Richards (1983)	Lecture, discussion, and debate were mentioned as three such "events."	Statements, dialogues, and minitalks are used.

Table 1 (Continued)

Variable	Source	Summary	Current TOEFL
Visual cues	O'Neill (1954)	Visual cues, such as lip movements, are used by some listeners to comprehend, especially when there is interfering noise.	No visual cues are provided on TOEFL.
Speaker-related			
Manner of delivery (skill of speaker)	Carroll (1971)	Manner of delivery effects comprehension, possibly by arousing greater attention.	Stimuli are delivered by skilled speakers.
Rate of speech	Foulke & Sticht (1969)	Intelligibility of speech is maintained up to about 275 words per minute for native speakers. Beyond this, intelligibility and comprehension suffer sharp losses.	Stimuli are presented slowly and deliberately--probably well below 275 wpm.
Rate of delivery, rhythm and stress, pauses and speech errors	Richards (1983)	These factors were viewed as ones that influence the work that listeners must do to process speech.	No speech errors, few if any pauses, pauses deliberate rate of delivery.
Rate of speaking/ redundancy, pauses	Flaherty (1979)	Rapid speech is more easily processed when it has redundancies or pauses.	Speech is not rapid on TOEFL. Little redundancy is apparent.
Degree of extended speech. Speed, slurring.	Kalivoda (1981)	Suggested that these are variables requiring attention in developing advanced listening comprehension in a foreign language.	Extended speech is not employed on TOEFL. No slurring or rapid speech is used.

Table 1 (Continued)

Variable	Source	Summary	Current TOEFL
Accent/dialect	Eisenstein & Berkowitz (1981)	Found that standard English was more comprehensible to adult ESL learners than were foreign accents or working class ("New Yorkese") English.	Standard American English is used.
	Wilcox (1978)	Comprehensibility varied with accent (but accent was confounded with rate of speech).	Standard American English is used.
Context-related			
Distractions during listening (competing stimuli)	Carroll (1971)	The more similar the competing message to the target message, the greater the interference.	No competing messages are used.
Attentional processes	Rubin (1982)	Listening performance is very dependent on (and moderated by) such factors as fatigue, motivation, antagonism towards the speaker, and disinterest in the topic.	- - - - - - - - -
Interval between listening and testing; note-taking	Yates (1979)	The contention is made that retention with and without note-taking is about the same up to 30 minutes; thereafter note-taking increases retention.	Note-taking is not permitted. Questions follow very closely after the stimulus.
Note-taking	Howe (1970) Weiland and Kingsbury (1979)	Students who took notes performed better on subsequent multiple-choice tests.	Note-taking is not permitted on TOEFL.

Table 2

Ratings of Importance for Academic Success of Listening, Reading, Speaking, and Writing Skills by Discipline

Skill		Discipline					
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22
Listening	M	4.27	4.38	4.36	4.39	4.19	4.18
	SD	.76	.79	.71	.64	.91	.89
Reading	M	4.73	4.86	4.82	4.83	4.76	4.68
	SD	.50	.35	.39	.38	.53	.55
Speaking	M	3.22	3.43	3.50	3.09	3.19	4.05
	SD	.87	.90	.94	1.02	1.10	.93
Writing	M	3.84	4.38	4.91	3.52	3.90	4.59
	SD	.89	.72	.29	.77	.61	.58

Note. Ratings were on a five-point scale ranging from 1 = low to 5 = high.

Table 3

Ratings of Importance of Specific Listening Activities by Discipline

Listening Activity		Importance						
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
1. Identifying the topic of a lecture.	M	4.35	4.65	4.73	4.48	4.52	4.14	4.46
	SD	.88	.48	.62	1.02	.66	1.01	.84
2. Identifying major ideas, themes, or theses in a lecture.	M	4.65	4.76	4.86	4.78	4.71	4.64	4.73
	SD	.53	.43	.46	.51	.63	.57	.53
3. Identifying supporting ideas and examples given in a lecture.	M	4.38	4.00	4.41	4.17	4.62	4.05	4.28
	SD	.59	.82	.72	.82	.72	.88	.77
4. Identifying relationships among major ideas, hypotheses, examples, etc., within a lecture.	M	4.46	4.76	4.64	4.43	4.57	4.29	4.52
	SD	.64	.43	.57	.71	.66	.63	.63
5. Identifying the role of discourse cues, e.g., words that signal such structural features of lectures as contrast ("but," "nevertheless"), addition ("furthermore"), and equivalence ("likewise").	M	3.59	3.52	4.23	3.65	3.52	2.86	3.57
	SD	.97	1.05	.90	.81	1.05	1.10	1.05
6. Inferring relationships or connections between information described in a lecture; understanding the distinction between causes and effects.	M	4.19	4.67	4.45	4.48	4.24	4.05	4.33
	SD	.80	.64	.72	.71	.75	.88	.79
7. Comprehending key vocabulary related to the topic of the lecture.	M	4.30	4.33	4.64	4.48	4.43	3.86	4.34
	SD	.65	.78	.57	.93	1.05	.87	.84
8. Deducing the meaning of words from the context in which they appear.	M	3.59	3.52	4.32	3.35	3.43	3.36	3.60
	SD	1.03	1.18	.70	1.00	1.14	.98	1.06
9. Recognizing the function of such features as change of pitch, sentence stress, and pauses in signaling the information structure of a lecture.	M	2.57	2.81	3.67	2.65	2.71	2.41	2.77
	SD	1.10	.91	1.04	1.00	1.24	.89	1.11

Table 3 (Continued)

Listening Activity		Importance						
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
10. Recognizing the lecturer's attitudinal signals, e.g., tone of voice, sarcasm, and humor.	M	2.30	2.67	3.67	2.26	2.48	2.43	2.59
	SD	1.04	.99	.89	1.03	.96	1.18	1.12
11. Following different modes of lectures.								
A. Spoken	M	4.03	4.67	4.67	4.67	3.86	4.32	4.33
	SD	1.22	.56	.64	.64	1.25	.76	.99
B. Audio	M	2.94	3.10	3.00	2.40	2.47	2.65	2.78
	SD	1.28	1.41	1.25	1.32	1.60	1.35	1.39
C. Audio/Visual	M	3.24	3.52	3.22	3.00	3.06	3.37	3.24
	SD	1.28	1.37	1.31	1.41	1.68	1.13	1.37
12. Following lectures given with different								
A. accents or dialects	M	3.27	2.45	3.32	2.70	3.45	3.05	3.06
	SD	1.22	1.32	1.22	1.16	1.32	1.30	1.30
B. speeds	M	3.17	3.25	3.63	3.09	3.15	3.14	3.22
	SD	1.30	1.22	1.09	1.25	1.42	1.06	1.25
13. Understanding different styles of lecturing.								
A. Formally structured	M	3.65	3.86	4.09	3.87	3.80	3.55	3.79
	SD	1.28	.83	.95	1.15	1.40	1.20	1.18
B. Unplanned, informally structured	M	3.29	3.81	4.18	2.96	3.47	3.77	3.55
	SD	1.27	.79	.94	1.20	1.35	1.13	1.20
14. Recognizing irrelevant matters; jokes, digressions, meanderings.	M	2.59	2.62	2.68	2.13	2.52	2.45	2.51
	SD	1.28	1.09	1.06	1.12	1.05	1.23	1.17
15. Recognizing the function of nonverbal signals, e.g., gestures, as markers of emphasis and attitude.	M	2.36	2.43	2.82	1.96	2.24	2.55	2.39
	SD	1.08	.95	1.03	1.04	1.06	1.16	1.09
16. Retaining information through note-taking.	M	4.54	4.67	4.55	4.70	4.20	4.00	4.46
	SD	.64	.47	.66	.46	.98	.74	.71
17. Retrieving information from notes.	M	4.65	4.62	4.48	4.74	4.35	3.59	4.43
	SD	.53	.58	.73	.53	1.01	1.11	.85

Table 4

Ratings of the Degree to Which Native Speaking and Nonnative Speaking Students Encounter Problems with Specific Listening Activities

Listening Activity		Degree of Problem						
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
1. Identifying the topic of a lecture.	Native	1.72 (.84)	1.71 (.76)	2.15 (.79)	1.52 (.77)	1.71 (.76)	1.95 (.93)	1.78 (.84)
	Nonnative	2.78 (1.03)	2.52 (1.26)	2.86 (1.06)	2.22 (1.06)	2.20 (1.08)	2.32 (1.14)	2.51 (1.13)
2. Identifying major ideas, themes, or theses in a lecture.	Native	2.31 (.94)	2.29 (.88)	2.45 (.74)	2.04 (.75)	1.95 (.72)	2.27 (.86)	2.22 (.85)
	Nonnative	3.13 (.85)	3.05 (1.13)	3.41 (.94)	2.74 (1.07)	2.52 (.79)	3.14 (.83)	3.01 (.98)
3. Identifying supporting ideas and examples given in a lecture.	Native	2.39 (.92)	1.95 (.84)	2.50 (.97)	2.22 (.78)	2.25 (.83)	2.14 (.77)	2.26 (.83)
	Nonnative	3.25 (.89)	2.90 (1.02)	3.41 (.83)	3.26 (1.03)	2.95 (.86)	3.30 (.64)	3.19 (.91)
4. Identifying relationships among major ideas, hypotheses, examples, etc. within a lecture.	Native	2.63 (.85)	2.81 (.96)	2.85 (.96)	2.70 (.91)	2.50 (.92)	2.62 (.90)	2.68 (.92)
	Nonnative	3.56 (.80)	3.38 (1.09)	3.64 (.77)	3.74 (.94)	3.10 (.89)	3.43 (1.05)	3.49 (.94)
5. Identifying the role of discourse cues, e.g., words that signal such structural features of lectures as contrast ("but," "nevertheless"), addition ("furthermore"), and equivalence ("likewise").	Native	2.22 (1.08)	1.71 (.98)	1.95 (.97)	2.04 (.81)	1.90 (.77)	1.95 (.95)	1.99 (.96)
	Nonnative	3.31 (1.10)	2.95 (1.00)	3.27 (1.09)	3.22 (1.10)	3.16 (1.14)	2.76 (.92)	3.13 (1.08)
6. Inferring relationships or connections between information described in a lecture; understanding the distinction between causes and effects.	Native	2.54 (.84)	2.81 (1.18)	2.58 (.94)	2.68 (.87)	2.55 (.92)	2.86 (.77)	2.66 (.93)
	Nonnative	3.34 (.92)	3.38 (1.13)	3.52 (1.14)	3.65 (1.00)	3.11 (1.02)	3.38 (.90)	3.40 (1.03)

Table 4 (Continued)

Listening Activity		Degree of Problem: Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
7. Comprehending key vocabulary related to the topic of a lecture.	Native	2.11 (.99)	2.29 (.93)	2.45 (.92)	2.65 (.96)	2.20 (.93)	2.50 (.99)	2.35 (.98)
	Nonnative	3.36 (.98)	3.43 (.90)	3.50 (.89)	3.57 (1.14)	3.05 (1.23)	3.50 (.99)	3.41 (1.03)
8. Deducing the meaning of words from the context in which they appear.	Native	2.31 (.88)	2.29 (.93)	2.45 (1.07)	2.22 (1.02)	2.10 (1.00)	2.41 (.78)	2.30 (.95)
	Nonnative	3.31 (.81)	3.38 (1.00)	3.55 (.78)	3.39 (1.24)	3.15 (.96)	3.57 (.95)	3.38 (.97)
9. Recognizing the function of such features as change of pitch, sentence stress, and pauses in signaling the information structure of a lecture.	Native	1.75 (.92)	1.71 (.76)	1.79 (1.05)	1.52 (.73)	1.47 (.50)	1.91 (1.00)	1.70 (.87)
	Nonnative	2.75 (1.26)	2.86 (.99)	3.24 (.97)	2.91 (1.21)	2.42 (1.23)	2.76 (.97)	2.82 (1.15)
10. Recognizing the lecturer's attitudinal signals, e.g., tone of voice, sarcasm, and humor.	Native	1.89 (.99)	1.90 (1.02)	2.15 (1.01)	2.00 (1.17)	1.80 (.81)	1.67 (.71)	1.90 (.98)
	Nonnative	2.97 (1.32)	3.14 (.94)	3.59 (.83)	3.17 (1.18)	3.48 (1.41)	2.95 (1.28)	2.75 (.79)
11. Following different modes of lectures.								
A. Spoken	Native	1.94 (.86)	1.90 (.92)	1.79 (.83)	2.14 (.89)	1.58 (.75)	2.14 (.89)	1.93 (.88)
	Nonnative	2.97 (1.11)	3.10 (1.02)	3.19 (.85)	3.52 (.85)	2.60 (1.02)	3.05 (.72)	3.06 (.99)
B. Audio	Native	1.82 (.82)	1.70 (.84)	2.00 (1.03)	1.86 (.89)	1.59 (.84)	1.68 (.65)	1.78 (.85)
	Nonnative	2.76 (.99)	2.90 (1.18)	3.24 (1.11)	3.05 (1.36)	2.71 (1.23)	2.75 (.99)	2.88 (1.15)
C. Audiovisual	Native	1.79 (.81)	1.67 (.84)	1.53 (.72)	1.90 (.87)	1.44 (.70)	1.74 (.71)	1.70 (.80)
	Nonnative	2.63 (1.11)	2.52 (.79)	2.71 (.82)	2.90 (1.15)	2.25 (.97)	2.70 (.95)	2.63 (1.01)

Table 4 (Continued)

Listening Activity		Degree of Problem						
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
12. Following lectures given with different								
A. accents or dialects	Native	2.58	2.16	2.56	2.65	2.26	2.63	2.50
		(1.09)	(1.27)	(1.12)	(.96)	(.91)	(1.15)	(1.10)
	Nonnative	3.11	2.85	3.78	3.48	3.11	3.55	3.29
		(1.12)	(.96)	(.85)	(1.06)	(1.37)	(1.08)	(1.13)
B. speeds	Native	2.24	2.30	2.94	2.65	1.84	2.36	2.37
		(1.09)	(1.10)	(1.11)	(1.05)	(.99)	(1.19)	(1.13)
	Nonnative	3.21	3.70	4.11	3.70	3.00	3.48	3.50
		(1.29)	(1.05)	(.85)	(1.08)	(1.41)	(.79)	(1.17)
13. Understanding different styles of lecturing.								
A. Formally structured	Native	1.86	2.00	1.85	2.00	1.89	2.05	1.93
		(.82)	(.98)	(.79)	(.90)	(.91)	(.92)	(.89)
	Nonnative	2.69	2.86	2.68	2.91	2.53	2.90	2.76
		(1.08)	(.94)	(.82)	(.88)	(1.04)	(1.06)	(.99)
B. Unplanned, informally structured	Native	2.23	2.38	2.30	2.23	2.00	2.76	2.31
		(.99)	(1.05)	(.84)	(1.08)	(.94)	(.97)	(1.01)
	Nonnative	3.20	3.43	3.50	3.30	3.00	3.55	3.33
		(1.14)	(1.05)	(1.20	(1.33)	(1.29)	(1.08)	(1.19)
14. Recognizing irrelevant matter: jokes, digressions, meanderings.	Native	1.80	1.71	1.80	1.73	1.55	1.82	1.74
		(.95)	(.88)	(1.08)	(1.01)	(.67)	(1.07)	(.96)
	Nonnative	3.29	3.05	3.27	3.26	2.65	3.00	3.11
		(1.11)	(.90)	(1.01)	(1.36)	(1.49)	(1.04)	(1.18)
15. Recognizing the function of nonverbal signals, e.g., gestures, as markers of emphasis and attitude.	Native	1.78	1.62	1.55	1.77	1.45	1.73	1.67
		(.85)	(.72)	(.86)	(.95)	(.67)	(.81)	(.83)
	Nonnative	2.58	2.57	2.77	2.32	2.10	2.77	2.53
		(1.09)	(1.22)	(1.17)	(1.06)	(1.18)	(1.04)	(1.15)

Table 4 (Continued)

Listening Activity		Degree of Problem						
		Engineering N = 37	Psychology N = 21	English N = 22	Chemistry N = 23	Computer Science N = 21	Business N = 22	Total N = 146
16. Retaining information through note-taking.	Native	2.75 (.98)	2.71 (.93)	2.95 (.83)	2.83 (.87)	2.42 (.99)	2.41 (.78)	2.69 (.93)
	Nonnative	3.17 (1.12)	3.57 (.85)	3.29 (1.03)	2.87 (.95)	2.79 (1.10)	3.23 (1.00)	3.15 (1.05)
17. Retrieving information from notes.	Native	2.72 (.77)	2.76 (1.02)	2.89 (.81)	2.87 (.99)	2.42 (1.09)	2.38 (.71)	2.68 (.92)
	Nonnative	3.25 (.95)	3.29 (1.03)	3.43 (1.00)	2.91 (.97)	2.84 (1.27)	3.05 (1.11)	3.14 (1.07)

Note. Ratings were on a five-point scale with minor/rare = 1, moderate/occasional = 3, major/frequent = 5.

Table 5

Descriptions of Typical Courses

Feature		Descriptiveness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Purposes of lectures								
To interpret material in an assigned text or outside reading	M	3.78	3.38	4.41	4.26	3.71	3.64	3.86
	SD	1.06	1.21	1.03	.74	1.12	1.07	1.10
To supplement an assigned text where needed	M	3.95	4.00	4.23	3.96	4.00	4.00	4.01
	SD	1.01	.93	.90	.86	.82	.95	.93
To present new material and concepts that are not related to existing texts	M	3.59	4.19	3.62	3.35	3.48	3.73	3.65
	SD	1.15	.85	1.33	1.24	1.26	.86	1.16
Role of lecturer								
Interactive--to lead, formulate, and summarize discussion	M	3.17	3.71	4.52	2.48	2.76	4.13	3.43
	SD	1.21	.93	.73	1.35	.97	.92	1.27
Noninteractive--to present a relatively substantial amount of information with a minimum of questions and exchange	M	3.32	3.55	2.41	3.91	3.90	2.55	3.28
	SD	1.27	1.12	1.30	1.10	.75	1.16	1.28
Features of lectures								
Frequent restatement of new or difficult information	M	3.43	3.48	4.00	3.65	3.67	3.32	3.58
	SD	1.00	.91	.90	1.09	.94	.82	.98
Frequent use of examples or anecdotes to illustrate concepts and to apply principles	M	3.84	3.90	4.50	3.61	4.29	4.27	4.04
	SD	.92	.81	.58	1.05	.82	.62	.88

Table 5 (Continued)

Feature		Descriptiveness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Frequent use of analogies or more familiar similar concepts	M	3.46	3.71	4.09	3.65	3.62	3.68	3.68
	SD	.95	.88	.95	1.05	.84	.87	.95
Primarily inductive in nature (use of examples to build up to more general concepts or principles, going from the concrete to the abstract)	M	3.22	3.19	3.59	3.22	3.86	3.14	3.35
	SD	1.28	.73	1.11	1.18	.89	1.06	1.11
Primarily deductive (going from general principles to specific examples)	M	3.67	3.43	3.14	4.00	3.00	3.14	3.43
	SD	.91	.79	.97	.88	1.02	1.06	1.00
Other								
Frequent use of audiovisual materials	M	2.65	2.29	2.14	2.48	2.48	3.05	2.53
	SD	1.12	1.03	.92	1.25	1.47	1.26	1.21
Frequent use of handouts, notes, etc.	M	3.54	2.71	4.09	3.26	3.67	3.59	3.49
	SD	1.00	.98	1.04	1.22	.94	1.15	1.13

Note. Ratings were on a five-point scale ranging from 5 = very descriptive to 1 = not at all descriptive.

Table 6

Ratings of Appropriateness of Various General Tasks for Evaluating Listening Skills of International Students

General Task		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Matching or distinguishing. Choosing a picture or written response that corresponds with what was heard.								
1. Pictorial response	M	3.46	2.75	3.05	3.83	2.85	2.45	3.12
	SD	1.13	1.30	1.00	1.27	1.35	1.41	1.32
2. Written response	M	3.86	3.55	4.00	4.17	3.40	4.14	3.87
	SD	1.07	1.28	1.07	1.05	1.36	1.22	1.20
Transferring. Transferring information into another form, e.g., listening to a discussion and developing a chart, etc. to represent it.	M	4.08	3.95	3.59	4.04	3.70	3.64	3.86
	SD	.88	.72	1.03	1.04	1.10	1.03	.99
Transcribing. Listening and writing down what was said, e.g., dictation.	M	3.68	3.76	3.91	4.39	3.45	3.27	3.74
	SD	1.01	.97	1.04	.82	1.16	1.39	1.13
Skimming. Listening for the overall topic of what was heard.	M	3.59	4.24	4.27	3.74	3.55	3.77	3.83
	SD	.97	.61	1.01	.90	1.28	.85	1.00
Scanning. Listening for selected, specific pieces of information, e.g., listening to a news broadcast to determine an election winner.	M	3.35	3.86	3.86	3.43	2.95	3.45	3.48
	SD	.96	1.08	1.01	1.21	1.24	.89	1.10

Table 6 (continued)

General Task		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Extending or constructing. Going beyond what was said, e.g., reconstructing a dialogue when lines are missing, or providing a conclusion to a story.	M	3.65	3.62	3.64	3.09	3.40	3.41	3.48
	SD	1.02	1.17	.93	1.47	1.20	1.30	1.20
Condensing. For example, reducing what is heard to an outline of main points.	M	4.27	4.57	4.64	4.26	3.85	4.14	4.29
	SD	.83	.66	.64	.79	1.01	.87	.85
Answering. Providing answers to questions involving:								
1. Recall of details	M	4.03	4.19	4.64	4.35	4.25	3.82	4.19
	SD	.97	.79	.57	.76	.77	.89	.86
2. Inferences and deductions	M	4.27	4.24	4.82	4.31	4.35	4.18	4.35
	SD	.76	.97	.39	1.02	.91	.72	.84
Predicting. Guessing or predicting outcomes, causes, etc., based on what was heard.	M	3.50	3.71	4.10	3.57	3.65	3.75	3.69
	SD	.83	1.03	1.02	1.38	1.01	1.09	1.07

Note. Ratings were on a five-point scale ranging from 5 = very appropriate to 1 = very inappropriate.

Table 7

Ratings of Appropriateness of Various Specific Tasks for Evaluating the Listening Skills of International Students

Specific Task		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Appropriate response questions. Questions that involve selecting the most logical answer from several available choices.	M	3.86	3.95	4.09	4.30	4.21	4.15	4.07
	SD	1.06	.95	1.00	.69	.83	1.11	.97
Paraphrase answer questions. Questions that involve selecting the best paraphrase of a statement.	M	3.97	4.38	4.23	4.30	4.20	4.05	4.17
	SD	1.15	.84	1.17	1.04	1.21	.98	1.09
Pictorially presented comprehension questions. Questions that involve selecting the most appropriate picture to represent what is said.	M	3.95	3.90	3.50	4.65	4.05	3.36	3.91
	SD	1.06	1.02	.94	.63	1.07	1.19	1.08
Conversation questions (dialogues). Questions that involve an answer to show comprehension of a conversation between two or more speakers.	M	3.95	4.14	4.50	3.96	3.95	3.64	4.01
	SD	.98	.99	.66	1.16	1.24	1.37	1.11
Number questions. Questions that involve the comprehension of numbers and numerical concepts.	M	4.49	4.29	4.14	4.83	4.30	4.27	4.40
	SD	.83	.98	.99	.38	1.10	.91	.91

Table 7 (continued)

Specific Task		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Aural cloze questions. Questions that involve filling in the deleted words in aurally presented material.								
Type I: Presentation of a short passage or lecture followed by questions keyed to the missing words.	M	3.38	3.14	3.43	3.35	3.05	3.27	3.28
	SD	1.15	1.04	1.26	1.46	1.20	1.42	1.26
Type II: Same as Type I (including blanks) except that multiple-choice questions are used.	M	3.59	3.67	3.77	3.70	3.60	3.48	3.63
	SD	1.13	.94	1.13	1.04	1.11	1.22	1.10
Note-taking outline fill in. Same as above except that after examinees listen to the the passage they fill in a partially completed outline and finally answer questions based on the outline and memory.	M	3.89	3.76	3.91	3.78	3.40	3.76	3.77
	SD	1.11	.92	1.08	1.02	1.32	1.23	1.13

Table 7 (continued)

Specific Task		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Note-taking multiple choice. Same as above except that questions are multiple choice.	M	3.70	4.20	3.86	3.70	3.35	3.55	3.72
	SD	1.04	.81	1.10	1.00	1.24	1.20	1.10
Dictation. Questions that involve recording as faithfully as possible what is said.								
Type I: A whole passage is read at a normal speed and examinees are instructed to listen, not to write.	M	3.14	3.38	3.50	3.18	3.32	2.95	3.23
	SD	1.12	1.13	1.02	1.30	1.38	1.26	1.21
Type II (dicto-comp): A passage is read three or four times while examinees listen carefully. After the last reading, examinees write the passage.	M	3.19	3.86	3.71	3.52	3.25	3.67	3.50
	SD	1.04	1.12	1.12	1.17	1.30	1.28	1.19
Adequate outline. Questions that involve listening to a passage and then choosing the most adequate of several alternative outlines.	M	3.78	3.81	3.38	3.83	3.50	3.86	3.71
	SD	.90	.91	1.09	1.01	1.25	1.14	1.05

Note. Ratings were on a five-point scale ranging from 5 = very appropriate to 1 = very inappropriate.

Table 8

Ratings of Appropriateness of Various Stimuli for Measuring Listening Comprehension of International Students

Stimulus		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
A professor giving a classroom lecture	M	4.81	4.90	4.68	4.87	4.90	4.59	4.79
	SD	.51	.29	.63	.34	.30	.65	.50
A librarian giving information, e.g., on how to use a new computerized card catalog	M	4.22	4.19	4.09	4.13	4.05	3.41	4.03
	SD	.84	.96	.85	.80	.97	.78	.91
A registrar explaining procedures, e.g., for adding/dropping courses	M	4.03	4.14	4.05	3.70	3.95	3.48	3.90
	SD	1.03	.99	.82	.86	.86	.79	.94
A foreign student advisor giving information/advice, e.g., on course selection	M	4.30	4.43	4.33	4.00	4.05	3.95	4.19
	SD	.87	.85	.89	.83	1.12	.98	.94
A laboratory or teaching assistant explaining, e.g., an experiment or homework assignment	M	4.57	4.67	4.68	4.70	4.80	3.68	4.52
	SD	.59	.71	.70	.55	.51	1.18	.81
A student health service nurse or doctor providing, e.g., information on medical care	M	3.81	3.81	4.27	3.83	3.60	3.05	3.74
	SD	1.04	1.01	.75	.92	1.07	1.07	1.04
A fellow student offering, e.g., advice on studying for an impending test	M	4.14	3.76	4.05	3.65	3.75	3.45	3.83
	SD	.66	.87	.88	.81	1.13	.99	.91

Note. Ratings were on a five-point scale ranging from 5 = very appropriate to 1 = very inappropriate.

Table 9

Ratings of Appropriateness of Various "Speech Events" for Evaluating the Listening Skills of International Students

Speech Event		Appropriateness						
		Engineering	Psychology	English	Chemistry	Computer Science	Business	Total
Informal conversations	M	3.89	4.05	3.55	3.87	3.76	3.57	3.79
	SD	.81	.65	.94	.95	.97	.95	.90
Classroom lectures	M	4.69	4.95	4.77	4.65	4.62	4.45	4.69
	SD	.57	.21	.52	.87	.95	.84	.71
Discussions	M	4.50	4.38	4.73	4.09	4.19	4.36	4.39
	SD	.60	.58	.45	.85	1.01	.88	.76
Laboratory instructions	M	4.50	4.52	4.27	4.74	4.19	2.95	4.23
	SD	.65	.66	.75	.53	1.05	1.30	1.02
Orally administered exams	M	4.19	3.86	3.95	3.65	3.43	3.38	3.79
	SD	.94	.89	.88	1.34	1.29	1.50	1.18
Interviews	M	4.17	3.67	4.00	3.78	3.52	3.68	3.84
	SD	.76	.64	.76	.72	1.22	.92	.88
Audiovisual presentations	M	3.49	3.62	3.41	3.26	3.19	3.23	3.38
	SD	.97	.79	.65	.99	1.00	.85	.90

Note. Ratings were on a five-point scale ranging from 5 = very appropriate to 1 = very inappropriate.

63

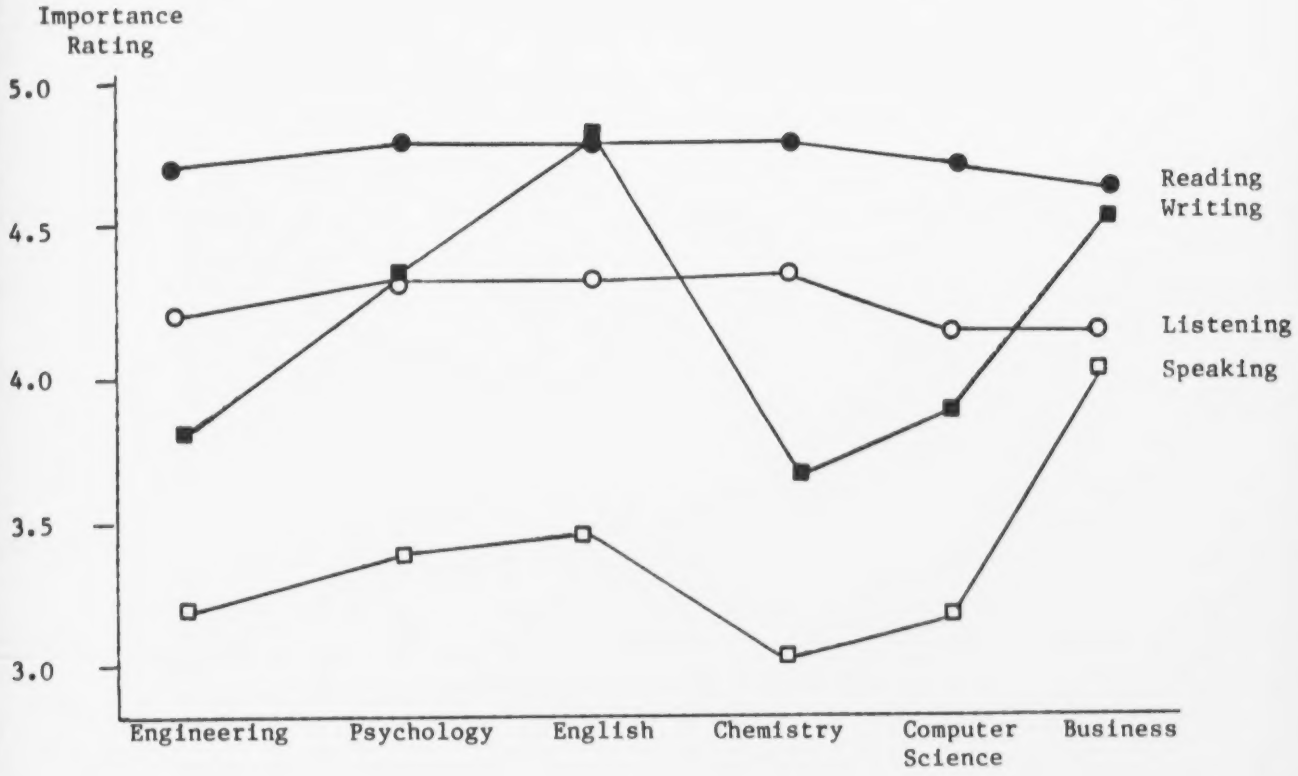

Figure 1. Ratings of Importance of Listening, Reading, Speaking, and Writing Skills

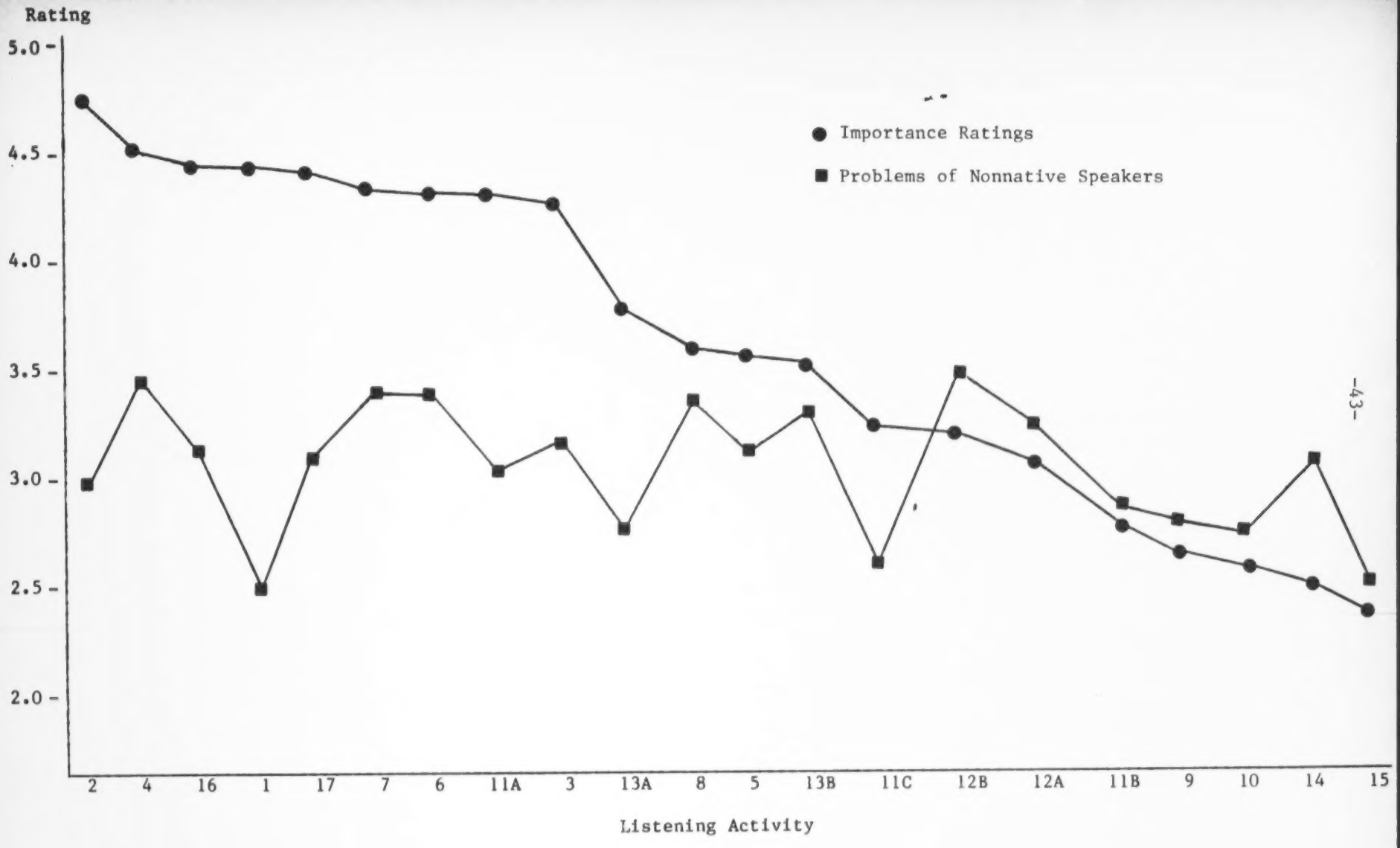

Figure 2. Ratings of Listening Activities: Importance and Problems for Nonnative Speakers

Note. See following page for complete description of activities.

1. Identifying the topic of a lecture.

2. Identifying major ideas, themes, or theses in a lecture.

3. Identifying supporting ideas and examples given in a lecture.

4. Identifying relationships among major ideas, hypotheses, examples, etc. within a lecture.

5. Identifying the role of discourse cues, e.g., words that signal such structural features of lectures as contrast ("but," "nevertheless"), addition ("furthermore"), and equivalence ("likewise").

6. Inferring relationships or connections between information described in a lecture; understanding the distinction between causes and effects.

7. Comprehending key vocabulary related to the topic of the lecture.

8. Deducing the meaning of words from the context in which they appear.

9. Recognizing the function of such features as change of pitch, sentence stress, and pauses in signaling the information structure of a lecture.

10. Recognizing the lecturer's attitudinal signals, e.g., tone of voice, sarcasm, and humor.

11. Following different modes of lectures.
    A. Spoken
    B. Audio
    C. Audiovisual

12. Following lectures given with different
    A. accents or dialects
    B. speeds

13. Understanding different styles of lecturing.
    A. Formally structured
    B. Unplanned, informally structured

14. Recognizing irrelevant matters; jokes, digressions, meanderings.

15. Recognizing the function of nonverbal signals, e.g., gestures, as markers of emphasis and attitude.

16. Retaining information through note-taking.

17. Retrieving information from notes.

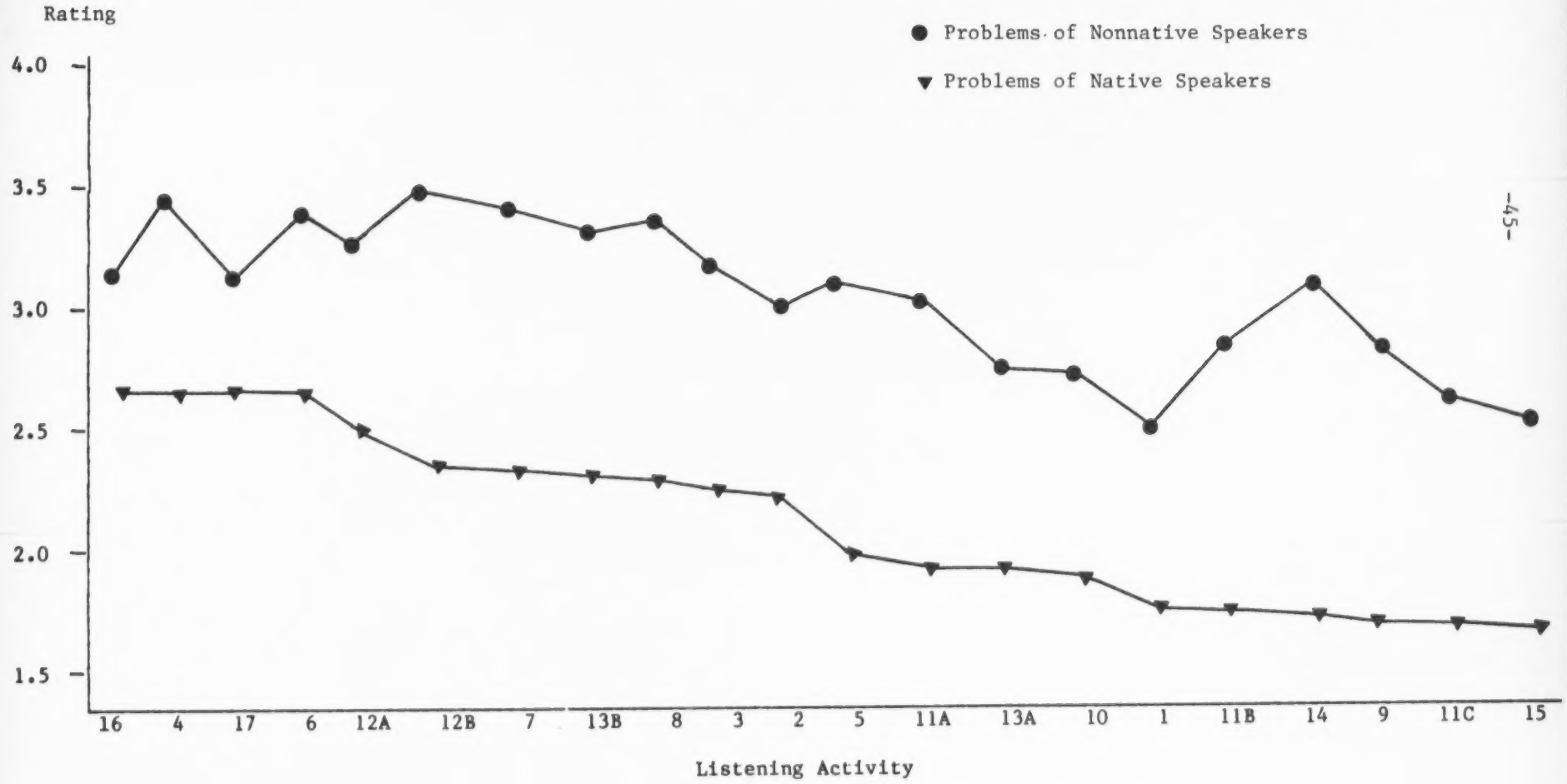

Figure 3. Ratings of Listening Activities: Problems of Native vs. Nonnative Speakers

Note. See following page for complete description of activities.

1. Identifying the topic of a lecture.

2. Identifying major ideas, themes, or theses in a lecture.

3. Identifying supporting ideas and examples given in a lecture.

4. Identifying relationships among major ideas, hypotheses, examples, etc. within a lecture.

5. Identifying the role of discourse cues, e.g., words that signal such structural features of lectures as contrast ("but," "nevertheless"), addition ("furthermore"), and equivalence ("likewise").

6. Inferring relationships or connections between information described in a lecture; understanding the distinction between causes and effects.

7. Comprehending key vocabulary related to the topic of the lecture.

8. Deducing the meaning of words from the context in which they appear.

9. Recognizing the function of such features as change of pitch, sentence stress, and pauses in signaling the information structure of a lecture.

10. Recognizing the lecturer's attitudinal signals, e.g., tone of voice, sarcasm, and humor.

11. Following different modes of lectures.
    A. Spoken
    B. Audio
    C. Audiovisual

12. Following lectures given with different
    A. accents or dialects
    B. speeds

13. Understanding different styles of lecturing.
    A. Formally structured
    B. Unplanned, informally structured

14. Recognizing irrelevant matters; jokes, digressions, meanderings.

15. Recognizing the function of nonverbal signals, e.g., gestures, as markers of emphasis and attitude.

16. Retaining information through note-taking.

17. Retrieving information from notes.

Trait	Method			
	TOEFL	Cloze	Condensing/Note-taking	Self-Ratings
Listening	X	X	X	X
A. Inferences/deductions	X			
B. Recall of details	X			
Reading	X	X	X	X
A. Inferences/deductions	X			
B. Recall of details	X			
Writing	X	--	--	X
A. Structure	X			
B. Written expression	X			
Speaking	--	--	--	X

Figure 4. Possible Configuration for Data Collection

APPENDIX

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08541

609-921-9000
CABLE-EDUCTESTSVC

DIVISION OF MEASUREMENT
RESEARCH AND SERVICES

June 26, 1984

Dear Faculty Member:

As part of a large research study, we are surveying faculty at 35 institutions to identify and describe the various listening skills that are necessary for successful academic performance in U.S. institutions. Our particular interest is in the first year of study for nonnative speakers of English. This research, which is being sponsored by the Test of English as a Foreign Language (TOEFL) program, is intended to provide a basis for further validation of the listening comprehension section of the TOEFL, an English language proficiency test that is used to facilitate admissions decisions about nonnative students.

We would appreciate your completing this questionnaire and returning it at your earliest convenience to the TOEFL research study coordinator at your school. Your responses will be treated confidentially. If you should have any questions please contact your research study coordinator.

As a small token of our appreciation we are able to pay you $20 if you provide the information requested on the enclosed payment form. Thank you in advance for your cooperation.

Sincerely,

Don Powers

Donald E. Powers
Project Director

DEP/ls
Enclosure

TOEFL SURVEY OF ACADEMIC LISTENING SKILLS

Please complete the following questions as best you can in order to help us determine (a) the most important listening skills required of first year students in your department and (b) the most appropriate ways of assessing these skills for the purpose of determining the admissibility of nonnative speaking students. You may find that some of the questions are more difficult to answer than others. We would, however, appreciate your best estimate even in those instances in which you may be unsure of your answer.

I. Background Information

Institution _____

Department _____

Number of students in your department* who are not native speakers of English. _____

Percentage of students in your department who are not native speakers of English. _____

Please indicate the geographical area from which most of your nonnative English speaking students come. (Circle only one.)

Africa. 1	Middle East. 5
Canada. 2	Asia (Far East). . . 6
Europe. 3	Other (specify). . . 7
Latin America . 4	_____

II. How important to success in your department are the following kinds of general abilities? (Please give an overall estimate, even though the importance of skills may depend on the type and level of courses as well as other factors.)

	Degree of Importance				
	Low		Moderate		High
A. Listening Skills	1	2	3	4	5
B. Reading Skills	1	2	3	4	5
C. Speaking Skills	1	2	3	4	5
D. Writing Skills	1	2	3	4	5

*If you are an undergraduate English faculty member, please give these numbers for all first year students at your institution, not just English majors.

III. For each of the following listening activities please indicate

(A) its importance to success in your department,

(B) the extent to which native speakers of English encounter problems with the activity, and

(C) the extent to which nonnative speakers of English encounter problems with the activity.

(Regarding (B) and (C), please base your responses on any information you feel is relevant. This might include, for example, your observations of students both in the classroom and outside it, students' performance on quizzes or exams, the questions students ask in class, and students' oral responses to questions posed in class.)

Activity	A Importance			B Problems of Native Speaking Students			C Problems of Nonnative Speaking Students		
	Low	Moderate	High	Minor/ Rare	Moderate/ Occasional	Major/ Frequent	Minor/ Rare	Moderate/ Occasional	Major/ Frequent
1. Identifying the topic of a lecture.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
2. Identifying major ideas, themes, or theses in a lecture.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
3. Identifying supporting ideas and examples given in a lecture.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
4. Identifying relationships among major ideas, hypotheses, examples, etc. within a lecture.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
5. Identifying the role of discourse cues, e.g., words that signal such structural features of lectures as contrast ("but," "nevertheless"), addition ("furthermore"), and equivalence ("likewise").	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
6. Inferring relationships or connections between information described in a lecture; understanding the distinction between causes and effects.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		
7. Comprehending key vocabulary related to the topic of a lecture.	1 2 3 4 5			1 2 3 4 5			1 2 3 4 5		

Activity	A Importance: Low		Moderate		High	B Problems of Native Speaking Students: Minor/ Rare		Moderate/ Occasional		Major/ Frequent	C Problems of Nonnative Speaking Students: Minor/ Rare		Moderate/ Occasional		Major/ Frequent
8. Deducing the meaning of words from the context in which they appear.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
9. Recognizing the function of such features as change of pitch, sentence stress, and pauses in signaling the information structure of a lecture.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
10. Recognizing the lecturer's attitudinal signals, e.g., tone of voice, sarcasm, and humor.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
11. Following different modes of lectures.															
A. Spoken	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
B. Audio	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
C. Audio/Visual	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
12. Following lectures given with different															
A. accents or dialects	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
B. speeds	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
13. Understanding different styles of lecturing.															
A. Formally structured	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
B. Unplanned, informally structured	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
14. Recognizing irrelevant matter: jokes, digressions, meanderings.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
15. Recognizing the function of non-verbal signals, e.g., gestures, as markers of emphasis and attitude.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
16. Retaining information through note-taking.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
17. Retrieving information from notes.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

Adapted from J. C. Richards (1983). Listening comprehension: Approach, design, procedure. TESOL Quarterly, 17, 219-240.

IV. How descriptive of the typical courses you teach is each of the following statements?

	Very descriptive		Somewhat descriptive		Not at all descriptive
A. Purposes of lectures					
1. To interpret material in an assigned text or outside reading	5	4	3	2	1
2. To supplement an assigned text where needed	5	4	3	2	1
3. To present new material and concepts that are not related to existing texts	5	4	3	2	1
4. Other (specify)_____ _____	5	4	3	2	1
B. Role of lecturer					
1. Interactive--to lead, formulate, and summarize discussion	5	4	3	2	1
2. Non-interactive--to present a relatively substantial amount of information with a minimum of questions and exchange	5	4	3	2	1
3. Other (specify)_____ _____	5	4	3	2	1
C. Features of lectures					
1. Frequent restatement of new or difficult information	5	4	3	2	1
2. Frequent use of examples or anecdotes to illustrate concepts and to apply principles	5	4	3	2	1
3. Frequent use of analogies or more familiar similar concepts	5	4	3	2	1
4. Primarily inductive in nature (use of examples to build up to more general concepts or principles, going from the concrete to the abstract)	5	4	3	2	1

5. Primarily deductive (going from general principles to specific examples)	5	4	3	2	1
6. Other (specify) _____ _____	5	4	3	2	1

D. Other

1. Frequent use of audio/ visual materials	5	4	3	2	1
2. Frequent use of handouts, notes, etc.	5	4	3	2	1
3. Other (specify) _____ _____	5	4	3	2	1

V. In your opinion, how appropriate are the following general kinds of tasks for evaluating the listening skills of international students? (Each kind of task could include a range of material from very simple to quite complex.)

Exercises involving:

	Very Appropriate (5)	Somewhat Appropriate (4)	Neither Appropriate nor Inappropriate (3)	Somewhat Inappropriate (2)	Very Inappropriate (1)
A. Matching or distinguishing. Choosing a picture or written response that corresponds with what was heard.					
1. Pictorial response	5	4	3	2	1
2. Written response	5	4	3	2	1
B. Transferring. Transferring information into another form, e.g., listening to a discussion and developing a chart, etc. to represent it.	5	4	3	2	1
C. Transcribing. Listening and writing down what was said, e.g., dictation.	5	4	3	2	1
D. Skimming. Listening for the overall topic of what was heard.	5	4	3	2	1
E. Scanning. Listening for selected, specific pieces of information, e.g., listening to a news broadcast to determine an election winner.	5	4	3	2	1

F. Extending or constructing. Going beyond what was said, e.g., reconstructing a dialogue when lines are missing or providing a conclusion to a story.	5	4	3	2	1
G. Condensing. For example, reducing what is heard to an outline of main points.	5	4	3	2	1
H. Answering. Providing answers to questions involving:					
1. Recall of details	5	4	3	2	1
2. Inferences and deductions	5	4	3	2	1
I. Predicting. Guessing or predicting outcomes, causes, etc., based on what was heard.	5	4	3	2	1

If you rated any of the exercises (A - I) above as inappropriate (ratings of 2 or 1), could you please explain briefly why you think they are inappropriate?

Exercise (Give letter)	Reason for inappropriateness

Note. Adapted from J. C. Richards (1983). Listening comprehension: Approach, design, procedure. TESOL Quarterly, 17, 219-240.

VI. How appropriate is each of the following specific kinds of questions for evaluating the listening skills of international students? (The material presented in each of these kinds of questions could be more or less difficult than that given here.)

Very Inappropriate (1)
Somewhat Inappropriate (2)
Neither Appropriate nor Inappropriate (3)
Somewhat Appropriate (4)
Very Appropriate (5)

A. Appropriate response questions. Questions that involve selecting the most logical answer from several available choices: For example, the examinee hears:
"When was Mrs. Stephens here?"
The examinee reads:
A. Tomorrow
B. Last year
C. To study
D. Her aunt

5 4 3 2 1

B. Paraphrase answer questions. Questions that involve selecting the best paraphrase of a statement. For example, the examinee hears:
"Jane became famous and wrote a book."
The examinee reads:
A. Her book made Jane famous.
B. After becoming famous, Jane became an author.
C. There's a famous book about Jane.
D. The book discusses Jane's fame. 5 4 3 2 1

C. Pictorially presented comprehension questions. Questions that involve selecting the most appropriate picture to represent what is said.
The examinee hears:
"Which picture shows the small square is in the upper left corner?"
The examinee sees:

5 4 3 2 1

(Questions of this type could also include pictures of people, places, etc.)

D. Conversation questions (dialogues). Questions that involve an answer to show comprehension of a conversation between two or more speakers.

The examinee hears:

Man: "I'm having trouble getting the results of my experiment to come out right."

Woman: "Why not try it this way and see what happens?"

Third voice: "What does the woman suggest?"

The examinee reads:

A. The experiment will never work.
B. The results will be a problem.
C. He should do the experiment differently.
D. He must not go out.

5 4 3 2 1

E. Number questions. Questions that involve the comprehension of numbers and numerical concepts. For example,

The examinee hears:

"The population of China is one quarter of the world's people."

The examinee reads:

A. China has four billion people.
B. The world's population is divided by four.
C. One-fourth of the world's people live in China.
D. China has the fourth largest population.

5 4 3 2 1

F. Aural cloze questions. Questions that involve filling in the deleted words in aurally-presented material.

Type I: Presentation of a short passage or lecture followed by multiple choice questions keyed on the missing words. The examinee hears a listening passage and is told:

"Listen to the tape and write down the missing words in the text."

Welcome to Physics One. This is an _____ course. Some of you _____ it Physics for Poets, so obviously you don't expect any previous _____. Those of you who have _____ any physics should leave now, and transfer to a _____ course. You'd just be _____ in this course. In lectures, I'll be covering all the basic _____ of physics. In lab you will have an _____ to see what I've been talking about.

5 4 3 2 1

Type II: Same as Type I (including blanks) except that multiple-choice questions are used. The examinee hears a listening passage and then chooses an answer from several alternatives.

The examinee hears:

"What does the speaker ask the students who have already studied physics to do?"

The examinee reads:

A. Enroll in another class.
B. Avoid talking.
C. Help answer questions.
D. Work with her for a semester.

5 4 3 2 1

G. Note-taking outline fill in. Same as Type F except that after examinees listen to the passage they fill in a partially completed outline and finally answer questions based on the outline and memory.

The examinee hears:
"Listen to the passage and fill in the missing words or phrases."
The examinee reads:
An introductory course in _____.
Not recommended for _____.
Course will cover _____.
The lab will be for _____.

Then examinee reads:
"Answer these questions based on your notes and memory."
1. Why shouldn't those with previous experience in physics take this course?

   _____

2. What will be discussed in the course?

   _____ 5 4 3 2 1

H. Note-taking multiple choice. Same as for G above except that questions are multiple choice. For example, "What does the speaker ask students who have already studied physics to do?"

A. Enroll in another class.
B. Avoid talking.
C. Help answer questions.
D. Work with her for a semester. 5 4 3 2 1

I. Dictation. Questions that involve recording as faithfully as possible what is said.

Type I. A whole passage is read at a normal speed and examinees are instructed to listen, not to write. The passage is reread a phrase at a time, with pauses, and examinees are required to write down each phrase. Finally the entire passage is reread at normal speed, during which examinees may make any final revisions. (Scoring is based on content accuracy and grammatical accuracy.) 5 4 3 2 1

Type II (dicto-comp): A passage is read three or four times while examinees listen carefully. After the last reading the examinees write the passage as they have heard it. Scoring is based primarily on content accuracy, i.e., the extent to which examinees recall the main points of the passage.

5 4 3 2 1

J. Adequate outline. Questions that involve listening to a passage and then choosing the most adequate of several alternative outlines.

5 4 3 2 1

If you rated any of the kinds of questions (A - J) above as inappropriate (ratings of 2 or 1), could you please explain briefly why you think they are inappropriate?

Kind of Question (Give letter) Reason for inappropriateness

VII. Listed are a number of persons/situations that nonnative students might encounter in an academic setting. Please rate the appropriateness of each as a stimulus for testing the listening comprehension of nonnative students seeking admission to your school/department.

	Very Appropriate (5)	Somewhat Appropriate (4)	Neither Appropriate nor Inappropriate (3)	Somewhat Inappropriate (2)	Very Inappropriate (1)
A. A professor giving a classroom lecture	5	4	3	2	1
B. A librarian giving information, e.g., on how to use a new computerized card catalog	5	4	3	2	1
C. A registrar explaining procedures, e.g., for adding/dropping courses	5	4	3	2	1
D. A foreign student advisor giving information/advice e.g., on course selection	5	4	3	2	1
E. A laboratory or teaching assistant explaining, e.g., an experiment or homework assignment	5	4	3	2	1
F. A student health service nurse or doctor providing, e.g., information on medical care	5	4	3	2	1
G. A fellow student offering, e.g., advice on studying for an impending test	5	4	3	2	1
H. Other (specify) _____ _____	5	4	3	2	1

Which one of the above do you regard as most appropriate? (Please select only one.) A B C D E F G H

Which one do you regard as least appropriate? (Please select only one.) A B C D E F G H

VIII. How would you rate the appropriateness of each of the following "speech events" as bases for assessing the listening skills of nonnative students? Please base your ratings on both the frequency and the importance of each activity.

Questions based on:	Very Appropriate (5)	Somewhat Appropriate (4)	Neither Appropriate nor Inappropriate (3)	Somewhat Inappropriate (2)	Very Inappropriate (1)
A. Informal conversations	5	4	3	2	1
B. Classroom lectures	5	4	3	2	1
C. Discussions	5	4	3	2	1
D. Laboratory instructions	5	4	3	2	1
E. Orally-administered exams	5	4	3	2	1
F. Interviews	5	4	3	2	1
G. Audio/visual presentations	5	4	3	2	1
H. Other (please specify) _____	5	4	3	2	1

Please describe briefly one or two "critical incidents" involving foreign students' understanding (or misunderstanding) of lectures, instructions, or other orally-presented material. What brought these incidents to your attention? Did they cause you to either raise or lower your estimation of a student's listening comprehension skills? Please explain. (For example, did a student badly misinterpret an assignment.)

_____

_____

_____

_____

_____

_____

_____

If you have any other comments you would like to make, we would welcome them.

_____

_____

_____

_____

_____

_____

May we tape-record one of your classroom lectures for use in further research? (We will not identify you or your institution in any subsequent research. Our interest is only in determining the content, style, and length of "typical" lectures in various academic fields.)

Yes _____ Your Name _____

No _____ Telephone number where we may contact you

( ) _____ _____

Thank you for your cooperation.

TOEFL Research Reports currently available . . .

Report 1. *The Performance of Native Speakers of English on the Test of English as a Foreign Language*. John L. D. Clark. November 1977.

Report 2. *An Evaluation of Alternative Item Formats for Testing English as a Foreign Language*. Lewis W. Pike. June 1979.

Report 3. *The Performance of Non-Native Speakers of English on TOEFL and Verbal Aptitude Tests*. Paul J. Angelis, Spencer S. Swinton, and William R. Cowell. October 1979.

Report 4. *An Exploration of Speaking Proficiency Measures in the TOEFL Context*. John L. D. Clark and Spencer S. Swinton. October 1979.

Report 5. *The Relationship between Scores on the Graduate Management Admission Test and the Test of English as a Foreign Language*. Donald E. Powers. December 1980.

Report 6. *Factor Analysis of the Test of English as a Foreign Language for Several Language Groups*. Spencer S. Swinton and Donald E. Powers. December 1980.

Report 7. *The Test of Spoken English as a Measure of Communicative Ability in English-Medium Instructional Settings*. John L. D. Clark and Spencer S. Swinton. December 1980.

Report 8. *Effects of Item Disclosure on TOEFL Performance*. Gordon A. Hale, Paul J. Angelis, and Lawrence A. Thibodeau. December 1980.

Report 9. *Item Performance Across Native Language Groups on the Test of English as a Foreign Language*. Donald L. Alderman and Paul W. Holland. August 1981.

Report 10. *Language Proficiency as a Moderator Variable in Testing Academic Aptitude*. Donald L. Alderman. November 1981.

Report 11. *A Comparative Analysis of TOEFL Examinee Characteristics, 1977-1979*. Kenneth M. Wilson. July 1982.

Report 12. *GMAT and GRE Aptitude Test Performance in Relation to Primary Language and Scores on TOEFL*. Kenneth M. Wilson. July 1982.

Report 13. *The Test of Spoken English as a Measure of Communicative Ability in the Health Professions: Validation and Standard Setting*. Donald E. Powers and Charles W. Stansfield. January 1983.

Report 14. *A Manual for Assessing Language Growth in Instructional Settings*. Spencer S. Swinton. February 1983.

Report 15. *Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*. Brent Bridgeman and Sybil Carlson. September 1983.

Report 16. *Summaries of Studies Involving the Test of English as a Foreign Language, 1963-1982*. Gordon A. Hale, Charles W. Stansfield, and Richard P. Duran. February 1984.

Report 17. *TOEFL from a Communicative Viewpoint on Language Proficiency: A Working Paper*. Richard P. Duran, Michael Canale, Joyce Penfield, Charles W. Stansfield, and Judith E. Liskin-Gasparro. February 1985.

Report 18. *A Preliminary Study of Raters for the Test of Spoken English*. Isaac I. Bejar. February 1985.

Report 19. *Relationship of Admission Test Scores to Writing Performance of Native and Nonnative Speakers of English*. Sybil B. Carlson, Brent Bridgeman, Roberta Camp, and Janet Waanders. August 1985.

Report 20. *A Survey of Academic Demands Related to Listening Skills*. Donald E. Powers. December 1985.

If you wish additional information about TOEFL research or would like to be placed on the mailing list to automatically receive order forms for newly published reports, write to:

TOEFL Program Office
CN 6155
Princeton, NJ 08541-6155
USA

57B612 • L125P5 • 275577 • Printed in U.S.A.